

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 03:52:00 : Search time 2120.21 Seconds
(without alignments)
60.819 Million cell updates/sec

Title: US-09-700-187-1
Perfect score: 12
Sequence: 1 ggatttaccagt 12

oring table: IDENTITY NUC
Gap 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	12	100.0	88	13	AZ925237
C 2	12	100.0	118	10	AW217731
C 3	12	100.0	126	13	AZ289196
C 4	12	100.0	130	10	AI610647
C 5	12	100.0	147	10	BE145737
C 6	12	100.0	147	11	D25785
C 7	12	100.0	150	10	AA579315
C 8	12	100.0	152	10	AU180504
C 9	12	100.0	155	13	AQ009485
C 10	12	100.0	165	10	AV333933
C 11	12	100.0	165	11	BF881282
C 12	12	100.0	171	11	BF909645

C 13	12	100.0	172	10	AI029683
C 14	12	100.0	173	13	AZ818519
C 15	12	100.0	177	13	AQ985182
C 16	12	100.0	180	11	BF235518
C 17	12	100.0	184	10	AV054690
C 18	12	100.0	185	10	AV071084
C 19	12	100.0	185	10	BB357868
C 20	12	100.0	185	11	BG942699
C 21	12	100.0	186	10	AI345298
C 22	12	100.0	189	11	F29452
C 23	12	100.0	190	10	AV288772
C 24	12	100.0	193	10	AV285178
C 25	12	100.0	194	10	AI136924
C 26	12	100.0	202	11	D60959
C 27	12	100.0	207	10	AI572468
C 28	12	100.0	210	10	AV336005
C 29	12	100.0	216	13	B25087
C 30	12	100.0	217	10	AI563911
C 31	12	100.0	220	13	AQ278868
C 32	12	100.0	223	10	BE236288
C 33	12	100.0	223	10	BE236293
C 34	12	100.0	224	10	AV252038
C 35	12	100.0	224	10	BB512144
C 36	12	100.0	225	10	BB256670
C 37	12	100.0	226	11	BI220740
C 38	12	100.0	227	10	AA129692
C 39	12	100.0	227	11	BI293223
C 40	12	100.0	228	10	BB171532
C 41	12	100.0	228	10	BB310621
C 42	12	100.0	230	10	BB152264
C 43	12	100.0	231	10	AV368688
C 44	12	100.0	231	10	BB503511
C 45	12	100.0	231	10	BB510596

ALIGNMENTS

RESULT 1
LOCUS AZ925237/6 88 bp DNA GSS 01-APR-2001
DEFINITION 4910.ez32kl8.sl Saccharomyces paradoxus N17 Saccharomyces paradoxus genomic clone 4910.ez32kl8.sl, DNA sequence.
ACCESSION AZ925237
VERSION AZ925237.1 GI:13496136
KEYWORDS GSS
SOURCE Saccharomyces paradoxus.
ORGANISM Saccharomyces paradoxus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 88)
AUTHORS Clifton,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish,W.R., Waterston,R.H. and Johnston,M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mjgenetics.wustl.edu
Class: random plasmid subclone.
Location/Qualifiers
FEATURES
source
1. 88
/organism="Saccharomyces paradoxus"
/strain="N17"
/db_xref="taxon:27291"
/clone="4910.ez32kl8.sl"
/clone_lib="Saccharomyces paradoxus N17"
/note="Random genomic sequence"
BASE COUNT 26 a 12 c 8 g 4 t

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 88;
 Best Local Similarity 100.0%; Pred. No. 7.3e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
 |||||
 Db 40 GGATTTTACAGT 29

RESULT 2

AW217731/c
 LOCUS AW217731 118 bp mRNA EST 18-MAY-2001
 DEFINITION EST296445 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cT0C6E5, mRNA sequence.

ACCESSION

AW217731 GI:6528605

VERSION

AW217731

KEYWORDS

EST.

SOURCE

tomato.

ORGANISM

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 118)

van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang

, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning

, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.

Generation of ESTs from tomato flower tissue

Unpublished (1999)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

Location/Qualifiers

1. .118

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone_lib="cT0C6E5"

/clone_lib="tomato flower buds 8 mm to pre-anthesis,

Cornell University"

/tissue_type="flower"

/dev_stage="buds 8mm-to-preanthesis"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XhoI; supplier: Tanksley; Flower buds and flowers were

taken from greenhouse plants (4-8 wks old, TA496). They

were immediately frozen in liquid nitrogen and then

size-separated while remaining frozen."

49 a 21 c 16 g 32 t

BASE COUNT 49 a 21 c 16 g 32 t

ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 118;

Best Local Similarity 100.0%; Pred. No. 7.2e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 47 GGATTTTACAGT 36

RESULT 3

AZ289196

LOCUS AZ289196 126 bp DNA GSS 27-JUL-2000

DEFINITION RPCI-23-59B23.TJB RPCI-23 Mus musculus genomic clone RPCI-23-59B23,

DNA sequence.

ACCESSION AZ289196

1 (bases 1 to 126)

Zhao, S., Nierman, W., Feidblyum, T., Malek, J., Shatsman, S., Akinret

, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 59 row: B column: 23

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .126

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-59B23"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACE3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies). "

38 a 20 c 30 g 38 t

BASE COUNT 38 a 20 c 30 g 38 t

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 126;

Best Local Similarity 100.0%; Pred. No. 7.1e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 93 GGATTTTACAGT 104

RESULT 4

AI610647/c

LOCUS AI610647 130 bp mRNA EST 21-APR-1999

DEFINITION tp20d02.xl NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2188323 3'

similar to SW:ULC6_HCMVA P16836 HYPOTHETICAL PROTEIN UL126.

; contains element LTR5 repetitive element ;, mRNA sequence.

ACCESSION AI610647

VERSION AI610647.1 GI:4619814

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 130)

REFERENCE 1 (bases 1 to 130)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Seq primer: -40up from Gibco

High quality sequence stop: 1.

FEATURES
source

Location/Qualifiers
1. .130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2188323"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT 36 a 21 c 28 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 130;
Best Local Similarity 100.0%; Pred. NO. 7.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 115 GGATTTTACAGT 104

RESULT 5

LOCUS BE145737 147 bp mRNA EST 21-JUN-2000
DEFINITION IL5-HT0207-231099-006-A01 HT0207 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE145737
VERSION BE145737.1 GI:8608461
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147)
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL
MEDLINE
COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=tl5-HT0207-231
099-006-A01&tl3=1999-10-23&tl4=1)

Seq primer: puc 18 forward

High quality sequence start: 56

High quality sequence stop: 147.

FEATURES
source

Location/Qualifiers
1. .147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0207"
/dev_stage="Adult"
/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 31 a 35 c 12 g 69 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 147;

Best Local Similarity 100.0%; Pred. NO. 7e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 132 GGATTTTACAGT 121

RESULT 6

LOCUS D25785/c 147 bp mRNA EST 30-NOV-1995
DEFINITION HUMG04153 Human colon mucosa Homo sapiens cDNA clone cm1984 3',
mRNA sequence.

ACCESSION D25785

VERSION D25785.1 GI:500469

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 147)

AUTHORS Okubo, K., Yoshii, J., Yokouchi, H., Kameyama, M. and Matsubara, K.

TITLE Global analysis of gene expression in colon mucosa: a large scale
random cDNA sequencing analysis

JOURNAL Unpublished (1994)

COMMENT

Contact: Okubo, K., Itoh, K., Yoshii, J., Yokouchi, H. and Matsubara, K.
Institute for Molecular and Cellular Biology
Osaka University
3-1 Yamada-oka, Suita, Osaka 565, Japan.

Location/Qualifiers

FEATURES
source

1. .147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cm1984"
/clone_lib="Human colon mucosa"
/note="Adult male, tissue_type = colon mucosa "

BASE COUNT 63 a 24 c 34 g 26 t
ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 147;

Best Local Similarity 100.0%; Pred. NO. 7e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 57 GGATTTTACAGT 46

RESULT 7
 AA579315 150 bp mRNA 12-SEP-1997
 LOCUS nf36e06.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:915874, mRNA
 DEFINITION sequence.
 ACCESSION AA579315
 VERSION AA579315.1 GI:2357499
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 150)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.,
 Michael Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 DNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
 Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 578 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 126.

FEATURES

Location/Qualifiers
 1..150
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:915874"
 /clone_lib="NCI_CGAP_Pr2"
 /sex="Male"
 /dev_stage="45 years old"
 /lab_host="DH10B"
 /note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
 strand cDNA was primed with oligo(dT)17 on 50 ng of
 DNase-treated, total cellular RNA obtained from 5,000-10
 ,000 microdissected preneoplastic cells
 histologically-determined to be prostatic intraepithelial
 neoplasia 2 (PIN2) cells. Double-stranded cDNA was
 ligated to EcoRI adaptors, 5 cycles of PCR applied to the
 cDNA with an adaptor-specific primer, and the resulting
 PCR product subcloned into pAMP10 by the UDG-cloning
 method (Life Technologies). Average insert size is 600
 bp. NOTE: Not directionally cloned. This library was
 constructed by David Krizman."
 54 a 20 c 30 g 46 t

BASE COUNT
 N
 Query Match 100.0%; Score 12; DB 10; Length 150;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggatttcacgt 12
 |||
 Db 126 GGATTTTACAGT 115

RESULT 8
 AU180504 152 bp mRNA 21-MAR-2001
 LOCUS AU180504 Medaka eye cDNA library (SNK01) Oryzias latipes cDNA clone
 DEFINITION NGX10.02c, mRNA sequence.
 ACCESSION AU180504
 VERSION AU180504.1 GI:13429341
 KEYWORDS EST.
 SOURCE Japanese medaka.

ORGANISM

Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 152)
 Sanaka,E., Hori,H., Naruse,K., Mitani,H. and Tanaka,M.
 Medaka EST analysis
 Unpublished (2001)
 Contact: Emi Sanaka
 Department of Biological Sciences
 Graduate School of Science, Nagoya University
 Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
 Tel: 81-52-789-2973
 Fax: 81-52-789-2974
 Email: sanaka@bio.nagoya-u.ac.jp
 This clone was isolated from Medaka eye cDNA library (SNK01) 5'end.

FEATURES

Location/Qualifiers
 1..152
 /organism="Oryzias latipes"
 /strain="wild type"
 /db_xref="taxon:8090"
 /clone="NGX10.02c"
 /clone_lib="Medaka eye cDNA library (SNK01)"
 /tissue_type="eye"
 /dev_stage="adult"
 /note="Wild samples from Okayama Pref. (Southern part of
 Japan)"
 56 a 23 c 30 g 43 t

BASE COUNT ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 152;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggatttcacgt 12
 |||
 Db 81 GGATTTTACAGT 92

RESULT 9

AQ009485 155 bp DNA 27-JUN-1998
 LOCUS CIT-HSP-2283N16.TRB CIT-HSP Homo sapiens genomic clone 2283N16, DNA
 DEFINITION sequence.
 ACCESSION AQ009485
 VERSION AQ009485.1 GI:3128870
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE AUTHORS

1 (bases 1 to 155)
 Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
 Simon,M. and Venter,J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building (1998)

TITLE JOURNAL COMMENT

Unpublished (1998)
 Other_GSSs: CIT-HSP-2283N16.TFB
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: M13 Reverse
 Class: BAC ends.


```

FEATURES             Location/Qualifiers
  source              1..155
                     /organism="Homo sapiens"
                     /db_xref="GDB:7148131"
                     /db_xref="taxon:9606"
                     /clone="2283N16"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT          30 a 46 c 34 g 45 t
ORIGIN
Query Match          100.0%; Score 12; DB 13; Length 155;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
      |||||
Db 11 GGATTTTACAGT 22

RESULT 10
AV333933
LOCUS
DEFINITION
MUS musculus cDNA clone 6330552F17 3', similar to X04070 Rat liver
mRNA for gap junction protein, mRNA sequence.
AV333933
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 165)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsura,S., Carninci,P., Akizawa,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
  source              1..165
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6330552F17"
                     /clone_lib="RIKEN full-length enriched, adult male medulla
                     oblongata"
                     /sex="male"
                     /tissue_type="medulla oblongata"
                     /dev_stage="adult"
                     /lab_host="DHI0B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGATTCTCGAGTTAATAATTAATCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified phuescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT          44 a 31 c 46 g 44 t
ORIGIN
Query Match          100.0%; Score 12; DB 10; Length 165;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
      |||||
Db 153 GGATTTTACAGT 164

RESULT 11
BF881282
LOCUS
DEFINITION
QV1-ET0183-021200-529-q05_1 ET0183 Homo sapiens cDNA, mRNA
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

```

Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QV1st2-QV1-ET0183-021200-529-d05-1st3-2000-12-02&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 165.

FEATURES

source

Location/Qualifiers

1. 165

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="ET0183"

/dev_stage="Adult"

/note="Organ: lung tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

34 a 46 c 35 g 50 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 165;
 -Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy; 1 ggattttacagt 12

|||||

Db 96 GGATTTTACAGT 107

RESULT 12

BF909645

LOCUS

BF909645 171 bp mRNA EST 18-JAN-2001

DEFINITION PM3-UT0058-181000-007-g01 UT0058 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF909645

VERSION BF909645.1

GI:12301103

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 171)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l1-PM3st2-PM3-UT0058-181000-007-g01st3-2000-10-18&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 17

High quality sequence stop: 113.

FEATURES

source

Location/Qualifiers

1. 171

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0058"

/dev_stage="Adult"

/note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

34 a 42 c 49 g 46 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 171;

-Best Local Similarity 100.0%; Pred. No. 7e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12

|||||

Db 117 GGATTTTACAGT 128

RESULT 13

AI029683/c

LOCUS

AI029683 172 bp mRNA EST 04-JUL-1999

DEFINITION UI-R-CO-iy-b-01-0-UI-sl UI-R-CO Rattus norvegicus cDNA clone

UI-R-CO-iy-b-01-0-UI 3', mRNA sequence.

ACCESSION AI029683

VERSION AI029683.1

GI:4300245

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

1 (bases 1 to 172)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

On Jun 22, 1998 this sequence version replaced gi:3247509.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult liver library. cDNA library preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1782730

Seq primer: M13 Forward

POLYA=No.

FEATURES

source

Location/Qualifiers

1. 172

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CO-iy-b-01-0-UI"

/clone_lib="UI-R-CO"

/dev_stage="adult"

/lab_host="DHI0B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CO library is a subtracted library derived from the UI-R-Al and UI-R-E1 libraries. The UI-R-Al library consisted of a mixture of individually tagged normalized libraries

constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-EI library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryos. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-EI clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-EI library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 32 a 47 c 32 g 61 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 29 GGATTTTACAGT 18

RESULT 14
A2818519 173 bp DNA GSS 20-FEB-2001
LOCUS
DEFINITION 2M0088A10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0088A10 R, DNA sequence.
ACCESSION A2818519
VERSION A2818519.1 GI:12988427
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 173)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0088 row: A column: 10
Seq primer: CACACAGGAACACTATGACC
Class: plasmid ends
High quality sequence stop: 173.
Location/Qualifiers
1. .173
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0088A10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 41 a 41 c 46 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 173;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||
Db 134 GGATTTTACAGT 145

RESULT 15
A0985182/c 177 bp DNA GSS 30-JAN-2000
LOCUS
DEFINITION RPCI-23-307M5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-307M5, DNA sequence.
ACCESSION A0985182
VERSION A0985182.1 GI:6818387
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 177)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-307M5.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
Plate: 307 row: M column: 5
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. .177
/organism="Mus musculus"
/strain="C57BL/6J"

```

/db_xref="taxon:10090"
/clone="RPCI-23-307M5"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "
```

```

BASE COUNT      56 a      48 c      20 g      52 t      1 others
ORIGIN

Query Match      100.0%; Score 12; DB 13; Length 177;
Best Local Similarity 100.0%; Pred. No. 7e+03; 0; Gaps 0;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggattttacagt 12
        |||
        139 GGATTTTACAGT 128
```

Search completed: April 3, 2002, 04:39:12
Job time: 2832 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 03:52:00 ; Search time 2120.21 Seconds
(without alignments)
60.819 Million cell updates/sec

Title: US-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estfun:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estom:*
- 5: em_estpl:*
- 6: em_estba:*
- 7: em_estro:*
- 8: em_estov:*
- 9: em_hic:*
- 10: gb_estl:*
- 11: gb_est2:*
- 12: gb_hic:*
- 13: gb_gss:*
- 14: em_gss_fun:*
- 15: em_gss_hum:*
- 16: em_gss_inv:*
- 17: em_gss_pln:*
- 18: em_gss_pro:*
- 19: em_gss_rod:*
- 20: em_gss_vrt:*
- 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	12	100.0	88	13	AZ925237
C 2	12	100.0	118	10	AW217731
C 3	12	100.0	126	13	AZ289196
C 4	12	100.0	130	10	A1610647
C 5	12	100.0	147	10	BE145737
C 6	12	100.0	147	11	D25785
C 7	12	100.0	150	10	AA579315
C 8	12	100.0	152	10	AU180504
C 9	12	100.0	155	13	AQ009485
C 10	12	100.0	165	10	AV333933
C 11	12	100.0	165	11	BF881282
C 12	12	100.0	171	11	BF909645

C 13	12	100.0	172	10	AI029683
C 14	12	100.0	173	13	AZ818519
C 15	12	100.0	177	13	AQ985182
C 16	12	100.0	180	11	BF235518
C 17	12	100.0	184	10	AV054690
C 18	12	100.0	185	10	AV071084
C 19	12	100.0	185	10	BB357868
C 20	12	100.0	185	11	BG942699
C 21	12	100.0	186	10	AI345298
C 22	12	100.0	189	11	F29452
C 23	12	100.0	190	10	AV288772
C 24	12	100.0	193	10	AV285178
C 25	12	100.0	194	10	AI136924
C 26	12	100.0	202	11	D60959
C 27	12	100.0	207	10	AI572468
C 28	12	100.0	210	10	AV336005
C 29	12	100.0	216	13	B25087
C 30	12	100.0	217	10	AI563911
C 31	12	100.0	220	13	AQ278868
C 32	12	100.0	223	10	BE236288
C 33	12	100.0	223	10	BE236293
C 34	12	100.0	224	10	AV252038
C 35	12	100.0	224	10	BB512144
C 36	12	100.0	225	10	BB256670
C 37	12	100.0	226	11	BI220740
C 38	12	100.0	227	10	AA129692
C 39	12	100.0	227	11	BI293223
C 40	12	100.0	228	10	BB171532
C 41	12	100.0	228	10	BB310621
C 42	12	100.0	230	10	BB152264
C 43	12	100.0	231	10	AV368688
C 44	12	100.0	231	10	BB503511
C 45	12	100.0	231	10	BB510596

ALIGNMENTS

RESULT 1
AZ925237/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AZ925237 88 bp DNA GSS 01-APR-2001
4910.ez32kl8.sl Saccharomyces paradoxus N17 Saccharomyces paradoxus
genomic clone 4910.ez32kl8.sl, DNA sequence.

AZ925237
A2925237.1 GI:13496136

GSS.
Saccharomyces paradoxus.

Saccharomyces paradoxus

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 88)

Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T., Gish

, W.R., Waterston, R.H. and Johnston, M.

Surveying Saccharomyces genomes to identify functional elements by

comparative DNA sequence analysis

Unpublished (2001)

Contact: Johnston M

Department of Genetics

Washington University Medical School

Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA

Tel: 314 362 2735

Fax: 314 362 7855

Email: mj@genetics.wustl.edu

Class: random plasmid subclone.

Location/Qualifiers

1. .88

/organism="Saccharomyces paradoxus"

/strain="N17"

/db_xref="taxon:27291"

/clone="4910.ez32kl8.sl"

/clone_lib="Saccharomyces paradoxus N17"

/note="Random genomic sequence"

26 a 12 c 8 g 42 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 88;
Best Local Similarity 100.0%; Pred. No. 7.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
DB 40 GGATTTTACAGT 29

RESULT 2

AW217731/c
LOCUS AW217731 118 bp mRNA EST 18-MAY-2001
DEFINITION EST296445 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cT0C6E5, mRNA sequence.

ACCESSION AW217731.1 GI:6528605
VERSION AW217731
KEYWORDS EST.
SOURCE tomato.

ORGANISM

Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE

1 (bases 1 to 118)
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
.F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
.C.M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato flower tissue
Unpublished (1999)

-TITLE

JOURNAL

COMMENT

Contact: CUGI
Clemson University
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES

source

Location/Qualifiers

1..118
/organism="Lycopersicon esculentum"
/cultivar="TA96"
/db_xref="taxon:4081"
/clone="cT0C6E5"
/clone_lib="tomato flower buds 8 mm to pre-anthesis,
Cornell University"
/tissue_type="flower"

/dev_stage="buds 8mm-to-preanthesis"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA96). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

BASE COUNT 49 a 21 c 16 g 32 t

ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 118;
Best Local Similarity 100.0%; Pred. No. 7.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
DB 47 GGATTTTACAGT 36

RESULT 3

AZ289196
LOCUS AZ289196 126 bp DNA GSS 27-JUL-2000
DEFINITION RPCI-23-59B23.TJB RPCI-23 Mus musculus genomic clone RPCI-23-59B23,
DNA sequence.
ACCESSION AZ289196

AZ289196.1 GI:9530982

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
1 (bases 1 to 126)
AUTHORS Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
.B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 59 row: B column: 23
Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..126
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-59B23"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

38 a 20 c 30 g 38 t

BASE COUNT 38 a 20 c 30 g 38 t

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 126;

Best Local Similarity 100.0%; Pred. No. 7.1e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||

DB 93 GGATTTTACAGT 104

RESULT 4

AI610647/c

LOCUS

AI610647 130 bp mRNA EST 21-APR-1999
DEFINITION tp20402.x1 NCI-CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2188323 3'
similar to SW:UIC6_HCMVA P16836 HYPOTHETICAL PROTEIN UL126.

:contains element LTR5 repetitive element ; , mRNA sequence.

ACCESSION AI610647.1 GI:4619814

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 130)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Seq primer: -400p from Gibco

High quality sequence stop: 1.

FEATURES

source

1. .130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CCGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 36 a 21 c 28 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 130;
Best Local Similarity 100.0%; Pred. No. 7.le+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12

Db 115 GGATTTTACAGT 104

RESULT 5

BE145737/c 147 bp mRNA EST 21-JUN-2000
LOCUS
DEFINITION ILS-HT0207-231099-006-A01 HT0207 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE145737
VERSION BE145737.1 GI:8608461
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 147)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,I.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5-HT0207-231099-006-A01&ts=1999-10-23&td=1)

Seq primer: puc 18 forward

High quality sequence start: 56

High quality sequence stop: 147.

FEATURES

source

1. .147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0207"
/dev_stage="Adult"
/note="Organ: head,neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No.196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 31 a 35 c 12 g 69 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 147;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12

Db 132 GGATTTTACAGT 121

RESULT 6

D25785/c 147 bp mRNA EST 30-NOV-1995
LOCUS
DEFINITION HUMGS04153 Human colon mucosa Homo sapiens cDNA clone cml984 3', mRNA sequence.

ACCESSION D25785

VERSION D25785.1 GI:500469

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 147)

AUTHORS Okubo,K., Yoshii,J., Yokouchi,H., Kameyama,M. and Matsubara,K.

TITLE Global analysis of gene expression in colon mucosa: a large scale random cDNA sequencing analysis

JOURNAL Unpublished (1994)

COMMENT Contact: Okubo,K., Itoh,K., Yoshii,J., Yokouchi,H. and Matsubara,K.

Institute for Molecular and Cellular Biology

Osaka University

3-1 Yamada-oka,Suita,Osaka 565,Japan.

Location/Qualifiers

source

1. .147

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="cml984"

/note="Adult male, tissue_type = colon mucosa"

BASE COUNT 63 a 24 c 34 g 26 t

ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 147;

Best Local Similarity 100.0%; Pred. No. 7e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12

Db 57 GGATTTTACAGT 46

```

RESULT 7
AA579315/c 150 bp mRNA 12-SEP-1997
LOCUS nf36e06.sl NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:915874, mRNA
DEFINITION sequence.
ACCESSION AA579315
VERSION AA579315.1 GI:2357499
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuacui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 578 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 126.
FEATURES
source
1..150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:915874"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."
BASE COUNT 54 a 20 c 30 g 46 t
ORIGIN
1 ggaatttcacgt 12
|||||
Db 126 GGATTTTACAGT 115

Query Match 100.0%; Score 12; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

RESULT 8
AA180504 152 bp mRNA 21-MAR-2001
LOCUS AU180504 Medaka eye cDNA library (SNK01) Oryzias latipes cDNA clone
DEFINITION NGV10.02c, mRNA sequence.
ACCESSION AU180504
VERSION AU180504.1 GI:13429341
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 152)
AUTHORS Sanaka,E., Hori,H., Naruse,K., Mitani,H. and Tanaka,M.
TITLE Medaka EST analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Emi Sanaka
Department of Biological Sciences
Graduate School of Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
Tel: 81-52-789-2973
Fax: 81-52-789-2974
Email: sanaka@bio.nagoya-u.ac.jp
This clone was isolated from Medaka eye cDNA library (SNK01) 5'end.
FEATURES
source
1..152
/organism="Oryzias latipes"
/strain="wild type"
/db_xref="taxon:8090"
/clone="NGV10.02c"
/clone_lib="Medaka eye cDNA library (SNK01)"
/tissue_type="eye"
/dev_stage="adult"
/note="Wild samples from Okayama Pref.(Southern part of
Japan)"
BASE COUNT 56 a 23 c 30 g 43 t
ORIGIN
1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

Query Match 100.0%; Score 12; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggaatttcacgt 12
|||||
Db 81 GGATTTTACAGT 92

RESULT 9
AA009485 155 bp DNA 27-JUN-1998
LOCUS CIT-HSP-2283N16.TRB CIT-HSP Homo sapiens genomic clone 2283N16, DNA
DEFINITION sequence.
ACCESSION AA009485
VERSION AA009485.1 GI:3128870
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2283N16.TFB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

```


constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryos. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 32 a 47 c 32 g 61 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
DB 29 GGATTTTACAGT 18

RESULT 14

AZ818519 173 bp DNA GSS 20-FEB-2001
LOCUS 2M0088A10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0088A10 R, DNA sequence.
ACCESSION AZ818519
VERSION A2818519.1 GI:12988427
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 173)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A., and Wright, D., Weiss, R.
REFERENCE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0088 row: A column: 10
Seq primer: CACACAGGAAACGCTATGACC
Class: plasmid ends
High quality sequence stop: 173.
Location/Qualifiers
1. 173
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0088A10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

FEATURES
SOURCE

1. 173
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0088A10"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 41 a 41 c 46 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 173;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
DB 134 GGATTTTACAGT 145

RESULT 15

AQ985182/c 177 bp DNA GSS 30-JAN-2000
LOCUS RPCI-23-307M5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-307M5, DNA sequence.
ACCESSION AQ985182
VERSION AQ985182.1 GI:6818387
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 177)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P., and Fraser, C.M.
REFERENCE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-307M5.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 307 row: M column: 5
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 177
/organism="Mus musculus"
/strain="C57BL/6J"

FEATURES
source

1. 177
/organism="Mus musculus"
/strain="C57BL/6J"

```
/db_xref="taxon:10090"  
/clone="RPCI-23-307M5"  
/clone_lib="RPCI-23"  
/sex="Female"  
/lab_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:  
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT      56 a      48 c      20 g      52 t      1 others  
-ORIGIN
```

```
Query Match      100.0%; Score 12; DB 13; Length 177;  
Best Local Similarity 100.0%; Pred. No. 7e+03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Uy 1 ggattttacagt 12  
    |||||  
Db 139 GGATTTTACAGT 128
```

Search completed: April 3, 2002, 04:39:12
Job time: 2832 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:09:50 : Search time 81.95 Seconds
(without alignments)
33.163 Million cell updates/sec

Title: US-09-700-187-1
Perfect score: 12
Sequence: 1 ggattttacagt 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	12	100.0	679	4	US-08-998-416-1071
C 2	12	100.0	1138	2	US-08-993-228-5
C 3	12	100.0	1434	2	US-08-903-801-2
C 4	12	100.0	1434	4	US-09-295-055-2
C 5	12	100.0	2259	1	US-07-828-700-7
C 6	12	100.0	3796	1	US-08-343-760A-1
C 7	11	91.7	38	1	US-08-061-697-33
C 8	11	91.7	38	1	US-08-061-697-34
C 9	11	91.7	38	1	US-08-131-365B-33
C 10	11	91.7	38	1	US-08-131-365B-34
C 11	11	91.7	38	2	US-08-668-123-33
C 12	11	91.7	38	2	US-08-668-123-34
C 13	11	91.7	332	1	US-07-748-761-1
C 14	11	91.7	492	1	US-08-470-720-3
C 15	11	91.7	856	4	US-09-318-661-3
C 16	11	91.7	856	4	US-09-318-661-6
C 17	11	91.7	926	4	US-09-230-380-9
C 18	11	91.7	935	1	US-08-162-475A-3
C 19	11	91.7	966	1	US-08-162-475A-1
C 20	11	91.7	1002	4	US-09-328-111-159
C 21	11	91.7	1076	2	US-08-934-959-1
C 22	11	91.7	1077	3	US-08-839-711-6
C 23	11	91.7	1195	1	US-08-181-271A-14
C 24	11	91.7	1195	1	US-08-449-315-14
C 25	11	91.7	1195	1	US-08-444-803-14
C 26	11	91.7	1195	1	US-08-449-043-14
C 27	11	91.7	1195	1	US-08-456-265A-14

28	11	91.7	1195	1	US-08-455-416-14	Sequence 14, Appl
29	11	91.7	1195	1	US-08-455-244-14	Sequence 14, Appl
30	11	91.7	1195	1	US-08-454-876-14	Sequence 14, Appl
31	11	91.7	1195	2	US-08-457-364-14	Sequence 14, Appl
32	11	91.7	1195	2	US-08-456-262-14	Sequence 14, Appl
33	11	91.7	1195	2	US-08-456-240-14	Sequence 14, Appl
34	11	91.7	1195	2	US-08-455-736-14	Sequence 14, Appl
35	11	91.7	1195	2	US-08-971-217-14	Sequence 14, Appl
36	11	91.7	1195	4	US-09-350-600-14	Sequence 14, Appl
C 37	11	91.7	1395	1	US-08-806-581A-1	Sequence 1, Appl
C 38	11	91.7	1569	2	US-08-743-637B-176	Sequence 176, App
C 39	11	91.7	1569	3	US-08-526-840B-176	Sequence 176, App
40	11	91.7	1569	3	US-08-821-984-9	Sequence 9, Appl
41	11	91.7	1569	4	US-09-329-749-9	Sequence 9, Appl
42	11	91.7	1613	1	US-08-219-842-1	Sequence 1, Appl
43	11	91.7	1613	1	US-08-451-096-1	Sequence 1, Appl
44	11	91.7	1613	2	US-08-810-599-1	Sequence 1, Appl
45	11	91.7	1622	2	US-09-014-969-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-998-416-1071/c
; Sequence 1071, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1071:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 679 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1651UP

CS
5/29/01

US-08-998-416-1071

Query Match 100.0%; Score 12; DB 4; Length 679;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 236 GGATTTTACAGT 225

RESULT 2

US-08-993-228-5/c
; Sequence 5, Application US/08993228
; Patent No. 5976838

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,228
; FILING DATE:
; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1138 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

US-08-993-228-5

Query Match 100.0%; Score 12; DB 2; Length 1138;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 985 GGATTTTACAGT 974

RESULT 3

US-08-903-801-2/c
; Sequence 2, Application US/08903801
; Patent No. 5932712

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/903,801
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0354 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNON01
; CLONE: 2272281
; US-08-903-801-2

Query Match 100.0%; Score 12; DB 2; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 1002 GGATTTTACAGT 991

RESULT 4

US-09-295-055-2/c
; Sequence 2, Application US/09295055
; Patent No. 6232440

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/295,055
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/903,801
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0354 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNON01
; CLONE: 2272281
; US-09-295-055-2

Query Match 100.0%; Score 12; DB 4; Length 1434;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
|||||

Db 1002 GGATTTTACAGT 991

RESULT 5

; Sequence 7, Application US/07828700

; Patent No. 5308760

; GENERAL INFORMATION:

; APPLICANT: Brown, Kit L.

; TITLE OF INVENTION: CRYSTAL PROTEINS OF BACILLUS

; TITLE OF INVENTION: THURINGIENSIS, GENES ENCODING THEM, AND HOSTS

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wegner, Cantor, Mueller & Player

; STREET: 1233 20th Street, N.W.

; CITY: Washington

; STATE: D.C.

; ZIP: 20036-8218

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/828,700

; FILING DATE: 19920203

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/817,915

; FILING DATE: 10 JAN 1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Cantor, Herbert I.

; REGISTRATION NUMBER: 24,392

; REFERENCE/DOCKET NUMBER: P-9575-23207

; TELECOMMUNICATION INFORMATION:

; US-09-295-055-2

; US-09-295-055-2

; US-09-295-055-2

; US-09-295-055-2

; US-09-295-055-2

; US-09-295-055-2

; US-09-295-055-2

; TELEPHONE: 202-887-0400
; TELEFAX: 202-835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2259 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: thompson1
; US-07-828-700-7

Query Match 100.0%; Score 12; DB 1; Length 2259;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12

|||||

Db 1536 GGATTTTACAGT 1547

RESULT 6

US-08-343-760A-1

; Sequence 1, Application US/08343760A

; Patent No. 5679783

; GENERAL INFORMATION:

; APPLICANT: De Robertis, Edward M

; APPLICANT: Sasal, Yoshiaki

; TITLE OF INVENTION: Tissue Differentiation Affecting

; TITLE OF INVENTION: Factor and Composition

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Majestic, Parsons, Siebert & Hsue

; STREET: Four Embarcadero Center, Suite 1450

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/343,760A

; FILING DATE: 22-NOV-1994

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Siebert, J. Suzanne

; REGISTRATION NUMBER: 28,758

; REFERENCE/DOCKET NUMBER: 3100.1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 363-5556

; TELEFAX: (415) 362-5418

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3796 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

US-08-343-760A-1

Query Match 100.0%; Score 12; DB 1; Length 3796;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12

Db 3249 GGATTTACAGT 3260
 |||||

RESULT 7
 US-08-061-697-33/c
 ; Sequence 33, Application US/08061697
 ; Patent No. 5498696
 ; GENERAL INFORMATION:
 ; APPLICANT: Brown, Michael S.; Briggs, Michael R.; Wang,
 ; APPLICANT: Xiaodong; Goldstein, Joseph L.
 ; TITLE OF INVENTION: Sterol Regulatory Element Binding Proteins
 ; TITLE OF INVENTION: and Their Use in Screening Assays
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/061,697
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: UTSD:347/PAR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 320-7200
 ; TELEFAX: (512) 474-7577
 ; INFORMATION FOR SEQ ID NO: 33:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 38 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-061-697-33

Query Match 91.7%; Score 11; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gattttacagt 12
 |||||

Db 27 GATTTTACAGT 17

RESULT 8
 US-08-061-697-34
 ; Sequence 34, Application US/08061697
 ; Patent No. 5498696
 ; GENERAL INFORMATION:
 ; APPLICANT: Brown, Michael S.; Briggs, Michael R.; Wang,
 ; APPLICANT: Xiaodong; Goldstein, Joseph L.
 ; TITLE OF INVENTION: Sterol Regulatory Element Binding Proteins
 ; TITLE OF INVENTION: and Their Use in Screening Assays
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/061,697
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: UTSD:347/PAR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 320-7200
 ; TELEFAX: (512) 474-7577
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 38 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-061-697-34

Query Match 91.7%; Score 11; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gattttacagt 12
 |||||

Db 16 GATTTTACAGT 26

RESULT 9
 US-08-131-365B-33/c
 ; Sequence 33, Application US/08131365B
 ; Patent No. 5527690
 ; GENERAL INFORMATION:
 ; APPLICANT: Brown, Michael S.
 ; APPLICANT: Briggs, Michael R.
 ; APPLICANT: Wang, Xiaodong
 ; APPLICANT: Goldstein, Joseph L.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
 ; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
 ; TITLE OF INVENTION: PROTEINS
 ; NUMBER OF SEQUENCES: 64
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: U.S.A.
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/131,365B
 ; FILING DATE: 01-OCT-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 418-3000
 ; TELEFAX: (512) 474-7577
 ; INFORMATION FOR SEQ ID NO: 33:
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-131-365B-33

Query Match 91.7%; Score 11; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
| | | | | | | | | |
Db 27 GATTTTACAGT 17

RESULT 10
US-08-131-365B-34
Sequence 34, Application US/08131365B
Patent No. 5527690
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TO STEROL REGULATORY ELEMENT BINDING
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/131,365B
FILING DATE: 01-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 474-7577
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-131-365B-34

Query Match 91.7%; Score 11; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
| | | | | | | | | |
Db 16 GATTTTACAGT 26

RESULT 11
US-08-668-123-33/c
Sequence 33, Application US/08668123
Patent No. 5891631
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TO STEROL REGULATORY ELEMENT BINDING
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/668,123
FILING DATE: 14-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/131,365
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-668-123-33

Query Match 91.7%; Score 11; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gattttacagt 12
| | | | | | | | | |
Db 27 GATTTTACAGT 17

RESULT 12
US-08-668-123-34
Sequence 34, Application US/08668123
Patent No. 5891631
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodong
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TO STEROL REGULATORY ELEMENT BINDING
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64

```

CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/668,123
FILING DATE: 14-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/131,365
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
- INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-668-123-34

Query Match 91.7%; Score 11; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 11; Conservative 0; Mismatches 0; Indels

QY 2 gattttacagt 12
|||||
DB 16 GATTTTACAGT 26

RESULT 13
US-07-748-761-1/c
Sequence 1, Application US/07748761
Patent No. 5304490
GENERAL INFORMATION:
APPLICANT: Grierson, Donald
APPLICANT: Fray, Rupert G
APPLICANT: Lycett, Grantley W
APPLICANT: Bird, Colin R
APPLICANT: Ray, John A
APPLICANT: Schuch, Wolfgang W
TITLE OF INVENTION: DNA, DNA constructs, cells and plants
TITLE OF INVENTION: derived therefrom
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Cushman, Darby and Cushman
STREET: Eleventh floor, 1615 L Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```

; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 101..218
; US-08-470-720-3

```

```

Query Match          91.7%; Score 11; DB 1; Length 492;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ggattttacag 11
    |||
Db 367 GGATTTTACAG 357

```

```

RESULT 15
US-09-318-661-3/c
; Sequence 3, Application US/09318661
; Patent No. 6268488
; GENERAL INFORMATION:
; APPLICANT: Barbas III, Carlos F.
; APPLICANT: Shabat, Doron
; APPLICANT: Rader, Christoph
; APPLICANT: List, Benjamin
; APPLICANT: Lerner, Richard A.
; TITLE OF INVENTION: PRODRUG ACTIVATION USING CATALYTIC ANTIBODIES
; FILE REFERENCE: PLF00115
; CURRENT APPLICATION NUMBER: US/09/318,661
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 856
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleotide
; OTHER INFORMATION: residue sequence of catalytic fragment.
; NAME/KEY: CDS
; LOCATION: (1)..(855)
US-09-318-661-3

```

```

Query Match          91.7%; Score 11; DB 4; Length 856;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 gattttacagt 12
    |||
Db 759 GATTTCAGT 749

```

```

Search completed: April 3, 2002, 05:14:51
Job time: 3901 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:14:06 ; Search time 1999.71 Seconds
(without alignments)
19180.752 Million cell updates/sec

Title: US-09-700-187-3

Perfect score: 2325

Sequence: 1 agccttaagcgaaggaa.....ttttgatcttgacaagaa 2325

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pi:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_om:*

20: em_or:*

21: em_ov:*

22: em_pat:*

23: em_ph:*

24: em_pi:*

25: em_ro:*

26: em_sts:*

27: em_sy:*

28: em_un:*

29: em_vi:*

30: em_htgo_hum:*

31: em_htgo_inv:*

32: em_htgo_rod:*

33: em_htg_hum:*

34: em_htg_inv:*

35: em_htg_rod:*

36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2325	100.0	3441	8	AB007911	AB007911 Pism sat
2	100	4.3	20785	8	AF138704	AF138704 Pism sat
3	89	3.8	5785	8	PSBLOX12	X78580 P.sativum (
4	82.4	3.5	326924	2	AC093082	Y17468 Pism sativ
5	82	3.5	1350	8	PSA17468	AC005083 Homo sapi
6	72.4	3.1	146285	9	AC005083	L06178 Apis mellif
7	72	3.1	16343	3	AMFGENOM	AC006970 Homo sapi
8	71.4	3.1	136098	9	AC006970	AC006279 Plasmodiu
9	68.6	3.0	245802	2	AC006279	AL592166 Homo sapi
10	68	2.9	160788	2	AL592166	AL593856 Homo sapi
11	67.8	2.9	156133	2	AL593856	AC072048 Mus muscu
12	67.6	2.9	208684	2	AC072048	AC005506 Plasmodiu
13	66.6	2.9	205429	2	AC005506	AL035477 Plasmodiu
14	66.6	2.9	224448	2	PFMALAP4	AC092304 Homo sapi
15	66.4	2.9	36977	2	AC092304	AE021553 Homo sapi
16	65.2	2.8	185691	2	AC021553	AE001398 Plasmodiu
17	64.4	2.8	14867	3	AE001398	AJ223323 Saccharom
18	64	2.8	1867	8	MTSCAJ23	AC011146 Homo sapi
19	63.8	2.7	158398	2	AC011146	AL158151 Human DNA
20	63.6	2.7	162445	9	AL158151	AX083744 Sequence
21	63.4	2.7	1141	6	AX083744	AC091925 Homo sapi
22	63	2.7	162914	2	AC091925	AL596253 Homo sapi
23	63	2.7	173349	2	AL596253	AL596253 Homo sapi
24	62.8	2.7	173349	2	AL596253	AC068363 Homo sapi
25	62.8	2.7	18529	2	AC068363	AC090493 Mus muscu
26	62.8	2.7	213530	2	AC090493	AC022281 Homo sapi
27	62.6	2.7	186431	2	AC022281	AF288092 Naegleria
28	62.4	2.7	193302	2	AC021992	M80833 Pea lox2-li
29	62.4	2.7	49843	3	AF288092	AL445219 Homo sapi
30	61.6	2.6	958	8	PEAL21LG	AC005969 Homo sapi
31	61.6	2.6	103649	2	AL445219	AL049180 Plasmodiu
32	61.4	2.6	147244	9	AC005969	AC012627 Homo sapi
33	61.4	2.6	192581	2	PFMAL13P1	AC026839 Homo sapi
34	61.2	2.6	100925	9	AC012627	AC021553 Homo sapi
35	61.2	2.6	156933	2	AC026839	AC008739 Homo sapi
36	61.2	2.6	185691	2	AC021553	AL031746 Plasmodiu
37	61.2	2.6	235532	9	AC008739	AP002820 Oryza sat
38	60.8	2.6	67970	3	PFMAL1P3	AL117204 Caenorhab
39	60.8	2.6	137332	8	AP002820	U49822 Saccharomyc
40	60.8	2.6	260341	3	CEY116A8C	M97514 Saccharomyc
41	60.6	2.6	2426	8	SDU49822	AC092304 Homo sapi
42	60.6	2.6	8622	8	YSCMTCYT0C	AC073269 Homo sapi
43	60.6	2.6	36977	2	AC092304	AL356793 Homo sapi
44	60.6	2.6	137889	9	AC073269	
45	60.2	2.6	140455	2	AL356793	

ALIGNMENTS

RESULT 1

AB007911

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Pism sativum

DNA

3441 bp

PLN

28-SEP-1999

complete cds.

GI:5926717

PRA2.

Pism sativum

DNA

Pism sativum

1 (sites)

Inaba,T., Nagano,Y., Sakakibara,T. and Sasaki,Y.

Identification of a cis-regulatory element involved in phytochrome

down-regulated expression of the pea small GTPase gene pra2

Plant Physiol. 120 (2), 491-500 (1999)

99292944

REFERENCE

2 (bases 1 to 3441)

AUTHORS Nagano,Y.
TITLE Direct Submission
JOURNAL Submitted (07-Oct-1997) Yukio Nagano, Nagoya University, Graduate School of Biological Sciences; Chikusa, Nagoya, Aichi 464-8601, Japan (E-mail:nagano@agr.nagoya-u.ac.jp, Tel:81-52-789-4168, Fax:81-52-789-4296)
FEATURES Location/Qualifiers
source 1..3441
/organism="Pisum sativum"
/db_xref="taxon:3888"
repeat_unit 904..1018
gene Join(2326..2567,2729..3188)
/gene="pra2"
CDS Join(2326..2567,2729..3188)
/gene="pra2"
/note="light-repressible GTP binding protein"
/codon_start=1
/product="Pra2"
/protein_id="BA84640.1"
/db_xref="GI:5926718"
/translation="MNQEMNGVEAEKLOEKIDYVFKVVVIGDSAVGKTOILSRFTKNE PCFDSKSLIGVEFOFKTYTINGKLLKAOIWDTPAGOEYRVAVTSAYIRGALGAMLYVDI TKRQFDHVAHWELRSHADGSIYIMLIGNKGLDYDQGVQTEDAVFEADOGLEFFS ETSFSGENVSFAFKLLQEIINKVSKRSLECNNGKNGHNDHVAALKEKIDIIISAS ELEISEIKLHSCSC"
BASE COUNT 1143 a 606 c 540 g 1152 t
ORIGIN
Query Match 100.0%; Score 2325; DB 8; Length 3441;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 aagctttaagcaaggaagcaacaattccaaaataataaaaactcctaaagaatga 60
DB 1 AGCCTTAAGGCAAGGAAAGACACAATTCAAAATAATATAAAACTCCTAAAGAAATGA 60
QY 61 tttattctattctataataaacttttctctatttccaaaaacacacatcaaaagtattgtga 120
DB 61 TTTTATTCTTATCTTCATAAATAACTTTTCTTATTCAAAACACATCAAAAGTTATGTGA 120
QY 121 ttcatacttttaattatcfcgataataataaattgtatattcaatttttatacaaatgtg 180
DB 121 TTCATATCTTTAATATCTGATAATAATAATTTGTTATTTCAATATTTTCATACAATTTG 180
QY 181 ttatatgaatatattgttagtgaaggaactaaagaataacctccgcaacatcaaaagtca 240
DB 181 TTATATGAATAATTGCTAGGTAAAGGCACTAAGANTAACTCCGCAACATCAAAAGTCA 240
QY 241 gaaacctctgttaactcttcagttgaaacgagaagaagtggaacacacagaaaaactaag 300
DB 241 GAAACCTCTGTAACTCTTCAGTTGAACGAGAAGGAAGTGGCAACACAGAAAACTAAG 300
QY 301 ttcccccacttaactctctgggttgggtgaggaactcccttcaattttatactctaaaga 360
DB 301 TTCCCCCACTTAACCTCTGTTGGTTGGGTGAGGACTTCCTTTTCAATTTATATCTTAAGA 360
QY 361 aatataatgacacttagatgggttgcatagctcatatatttttaagtaataatcacc 420
DB 361 AATACATTAGACACTAGATGGGTGCAATAGCTCATATATTTTAAAGTAATAATACCC 420
QY 421 acttcaagttttttgtttttgtgtgtgtagtagatgataagatggatcatctttctcaa 480
DB 421 ACTTCAAGTTTGTGTTTGTGTTGTCAGTAGATGATAAGATGGATCATTTCTCAA 480
QY 481 ggcctcttgaaaagacataaagatccataactccaccagaagattgctttacatcctaacca 540
DB 481 GGCCCTTATGCAAGACATAAAGATCCATATACTCCACCAGAGTTGCTTTACATCTTAACA 540
QY 541 agttaatgaattaaattctcgaacaataattttctaccagaaggaagtatttatatgcac 600
DB 541 AGTAAATGAATTAATTTCTTGAACAATAATTTCTTACCAAGGAAGTTTATATATGCAC 600

QY 601 attttcaatgtattttttatatagaattgatacatgtttctgtttatacaagattagaatt 660
DB 601 ATTUUCCTAATGTATTTTATATAGAATTGATACATGCTTTCTGTGTATACAAAGATTAGAATT 660
QY 661 tggatttctatcccaaacctcctacacttggtgagaaatttccagctccaacctcagtaaat 720
DB 661 TGGATTCTCTATCCAAACCTCCTACACTTGGTGAGAAATTTGACGCTCAACCTCAGTAAAT 720
QY 721 cagggttccctccctcaaacctcacaacttggttgagtgagaatttatgacgtcaacctagc 780
DB 721 CAGGTTCTCTCCCTTCAACCTCATACACTTGGTTGAGTGAGAAATTTATGAGCTCAACCTAG 780
QY 781 aatatgaatccctctcccaagatcctacacttatctgagtgagaatttttgctcctcgacct 840
DB 781 AATATGAATCCCTCTCCAAGATCCTACACTTATCTGAGTGAGAAATTTTGGTCTCTCGACCT 840
QY 841 caacaagatagatttgatgggttcacacagagggggaagcattcacattgggttcaaaatttc 900
DB 841 CAACAAGATAGATTGTGATGGGTCTCACAGGGGGAAGCATTCACATTTGGTCAAAAGATTTC 900
QY 901 acccaaacaaagtgagagagacatcacatatcaacacaaaaccttaaggtgtataggtgtatg 960
DB 901 ACCCAAAACAAAGTGAGAGACATCATATATCAACCAAAACCTTTAAGCTGATAGTGTATG 960
QY 961 agttctcttacttataaagtgtctcaacctccacttttctaaagcaatgtgtgacctagaac 1020
DB 961 AGTTCTCTTACTTATAAAGTGCTCAACCTCCACTTTTCTTAAGCAATGTGTGACTTAGAAC 1020
QY 1021 tcacacttattctcaacaataactcaacacttggtttatcaacaatctccccacacaagtgtg 1080
DB 1021 TCACACTTATTCTCAACATAACTCACACTTGTGTATCAACAATCTCCCCACAAAGTGTG 1080
QY 1081 agtcaattcgtctatgtccctccctcaagtgaattctcttcatccgcatgcttataccgttg 1140
DB 1081 AGTTCAATCGCTATGTCCCTCCCTCAAGTGAATCTCTTTTCATCCGCAATGTTATACCGT 1140
QY 1141 ttgacatacatatttactcgtcatgggcaacttcaatgggacacgctgcctgaccacacatg 1200
DB 1141 TTGACATACATCTTTACTCGTCATGGGCATTCAAATGGGACACGCTGCCTGACCACCATG 1200
QY 1201 tcaagaagacttttgacacaaaggagtcggtcccttactgaaccaggaactctgataccatt 1260
DB 1201 TCAAGAAGACTTTTGACACAAGAGTCGGTCCCTTACTCGAACACACACTCTGTATACCAT 1260
QY 1261 aatagatcactttgaatgatattcattactatcatatcaaacatttaccgttaaagataaaa 1320
DB 1261 AATAGATCACTTTGAATGGATATCATCTACTATATCAACACATTTACGTAAGATATAAA 1320
QY 1321 aattcaccacaaacaaatgagagagacactacatctctcttattatataataaaatgttaa 1380
DB 1321 AATTCAACCAACAAATGAGAGAGACACTACATCTCTCTTATATATTAATAAAATGTAA 1380
QY 1381 agaaaaataagtaataaaagtaaacacataattttgataaaatttattactaaaactatttcc 1440
DB 1381 AGAAAAATATAGTATAAAAGTAACACATATTTTGATAAAATTTTATTACTAAAACTATT 1440
QY 1441 tagtacctgttaactcatgctgaggttttaccagtaataaagaaacgaggtagcccaaac 1500
DB 1441 TAGTACTTGTGTAATCATGTCTGAGGATTTTACAGTAATAAAGNAACGAGGTAGCCCAAC 1500
QY 1501 aaaagtataaattgtgaggggtgtgactttgtcgggtgcaaaaaaagaacccccaaactt 1560
DB 1501 AAAAGTGATAAATTGTGGAGGTTGTGATCTTTGTGGTGTCAAAAAATGAAACCCCAAACTT 1560
QY 1561 gfgatatgtgcgactgctccgtcgctacattgaaatttaagtaattgtctttttataacg 1620
DB 1561 GTGATATTGTGCGACTGCTCCGTCGCTACATTTGAAATTAATGAATGTCTTTTATAACG 1620
QY 1621 ttgtctatgcccgtattaccctatattgctcactagaaatgggacaattgaatttaatatat 1680
DB 1621 TTTGTCTATGCCGTATTACCCATATGCTCAGTAGAATGGGACANCTGAATTTAATATATAT 1680
QY 1681 ctgtcatgtgtgggtgatttcaatttaattgtatcgttaattggtaggacatactcgtct 1740

```

Db 1681 CTGTCATGTGGTGGATTCAATTTAATGTATCGTAAATGGTAGGACATACATCGCT 1740
Qy 1741 acaaatatcatcatcactgctcaatcactggtcaatggtgtttctctctccatgaattc 1800
Db 1741 ACACAAATATATATCACTGCTCAATCAGTGGTCAATGTTCTCTCTCCCATGAATTC 1800
Qy 1801 acattgtaagaataattaccaccttaaaatgtttatcccttgacacatcttcacataa 1860
Db 1801 ACATTGCTAAAGAAATTAACCACTTAAATGTTATTCCTTGCACACATTCACATCA 1860
Qy 1861 ttattaaacattttaccatttgaaacacacatcatattcaatcaattatttttgatt 1920
Db 1861 TTTATTAAAAATTTTACCAATGGAACACACATACATATTCAATCAATATTATTGCA 1920
Qy 1921 ttcaaaactaaacaaacactagaatttttgaattttgaattatagcaaattttcaaaa 1980
Db 1921 TTCAAAAACCTAAACCAACAACTTAGAATATTGTAATTTATGACACAAATTTTCAA 1980
Qy 1981 tatcctagtcttcacactcaataattcacaatttcacaatcccttgcaaaaacatcac 2040
Db 1981 TATCCTAGTCTTCAACCACTCAATAATTCACAATTTCCAAATCCCTTGCAAAACATCA 2040
Qy 2041 acctctagaactttgatttaataatcctaataaaagcaataatgatgatctataaaca 2100
Db 2041 ACCCTAGAAACTTTGATTAAATATCTAATAATCTAATAAAGCAATAATATGATAT 2100
Qy 2101 caccatatgttatgataatatgatgcagcaatacacacttaatttggttaaacattaa 2160
Db 2101 CACCATATATGTTATGATATAATATGATGACGCAATACACTTAATTTGGTAAAGCAT 2160
Qy 2161 agcgagacaacttatlaacaccggttaattcaacaacccgttggtgagttcatgttt 2220
Db 2161 AGCGAGACAACCTATTAAACACCGTAATTCAACACCGTTGTTGTCAGTTTCATG 2220
Qy 2221 cttccaaactcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2280
Db 2221 CTTCCAACTCTTTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2280
Qy 2281 atatatactctcttgaacctcttcttcttcttcttcttcttcttcttcttcttct 2325
Db 2281 ATATATACTATCTCTCTGGAACCHCTTTTGTGATCTTGACAAGAA 2325

RESULT 2
AF138704 20785 bp DNA PLN 03-FEB-2000
LOCUS Pisum sativum gibberellin c20-oxidase gene, complete cds.
DEFINITION AF138704
ACCESSION AF138704
VERSION AF138704.1 GI:6855710
KEYWORDS
SOURCE pea.
ORGANISM Pisum sativum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
Pisum.
REFERENCE
1 (bases 1 to 20785)
Martin,D.N., Proebsting,W.M., Parks,T.D., Dougherty,W.G., Lange,T.,
Lewis,M.J., Gaskin,P. and Hedden,P.
Feed-back regulation of gibberellin biosynthesis and gene
expression in Pisum sativum L
Planta 200 (2), 159-166 (1996)
JOURNAL
MEDLINE
79060758
REFERENCE
2 (bases 1 to 20785)
Martin,D.N. and Proebsting,W.M.
Pisum sativum gibberellin c20-oxidase, genomic clone
Unpublished
JOURNAL
REFERENCE
3 (bases 1 to 20785)
Martin,D.N. and Proebsting,W.M.
Direct Submission
TITLE Submitted (25-MAR-1999) Horticulture, Oregon State University, ALS
4017, Corvallis, OR 97331-7304, USA

```

```

FEATURES
source
Location/Qualifiers
1..20785
/organism="Pisum sativum"
/strain="Alaska"
/db_xref="taxon:3888"
join(<1912..2468,2821..3142,3416..3880)
/product="gibberellin c20-oxidase"
join(1912..2468,2821..3142,3416..3679)
/function="converts GA53 to GA44 to GA19 to GA17 and GA20"
/note="converts GA12 to GA15 to GA24 to GA9 and GA25"
/note="oxoglutarate-dependent dioxygenase; similar to the
product encoded by GenBank Accession Number X91658"
/codon_start=1
/product="gibberellin c20-oxidase"
/protein_id="AAF29605.1"
/db_xref="GI:6855711"
/translation="MAECITSSAKLMTQKSDKNEGSSKLVFDASFLNQLNLPKQ
FIWPDDEKPCMNVPDLIDFKNLSGDPFAAMEASKTIGEACEKHGFFLVVNHGI
DTKLIEHAHSYMDNFPEVPLSQRCQRCQKGYASFTGRFSSNLPKLTLSQCF
SDKNSNIIVKDYLSNTLGEDFOFGVEYQYCEAMSKLSLIGIMELGMSLGVKECF
RDFEENKSIIMRLNYPYPCOKPDLTGLTGHPCDPTSLTILHODVGGLOVFDNEMHS
IRDNFAFVNVIGDTPMALNGRYKSLHRAVNVNKTTRKSLAFELCPKGDVVSPPS
ELVNDLTPRIYPPDFTWPMLEFETQKHVRADMRILEAFTKWIQQQN"
3'UTR 3680..3880
misc_feature 13001..13800
/note="similar to tobacco retroelement presented in
GenBank Accession Number X13777"
misc_feature 14272..14582
/note="similar to region upstream of pea chalcone synthase
gene presented in GenBank Accession Number X80007"
repeat_region 15844..15905
/rpt_type=tandem
/rpt_unit-at
BASE COUNT 6770 a 3474 g 7431 t 31 others
ORIGIN

Query Match 4.3%; Score 100; DB 8; Length 20785;
Best Local Similarity 84.8%; Pred. No. 1.8e-07;
Matches 112; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 963 ttctcttactataaagtctcaacctccacttttcttaagcaatgtgtgacttagaactc 1022
Db 3 TCCTCTAACTTATAAAGTGTTCACATCCACTTTTCTAAGCAATGTGAGACTTAGAATC 62

Qy 1023 acacttattctcaataactcaactgttttatcaacaatctccccaagtgtgag 1082
Db 63 ACACCTGTGTTCTTACGACAAATTCACATTTGTTTCTTAATATTTCTCCCTGAA 122

Qy 1083 ttcatcgcctat 1094
Db 123 TACATCCACTAT 134

RESULT 3
PSBLOX12
LOCUS PSBLOX12 5785 bp DNA PLN 06-APR-1995
DEFINITION P.sativum (Birte) Loxl:Ps:2 gene.
ACCESSION X78580
VERSION X78580.1 GI:469153
KEYWORDS lipoxigenase; Loxl gene.
SOURCE pea.
ORGANISM Pisum sativum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
Pisum.
REFERENCE
1 (bases 1 to 5785)
Casey,R.
Direct Submission
TITLE Submitted (01-APR-1994) R. Casey, John Innes Institute, Colney
Lane, Norwich, NR4 7UH, UK
JOURNAL
REFERENCE
2 (bases 1 to 5785)

```


University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

```
----- Project Information -----
Center project name: H_NH0505B03
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence,
* as soon as it is available and the accession number will
* be preserved.

1568:	contig of 1568 bp in length
1668:	gap of unknown length
1669:	contig of 1677 bp in length
3345:	gap of unknown length
3445:	gap of unknown length
4925:	contig of 1480 bp in length
5025:	gap of unknown length
7252:	contig of 2227 bp in length
7352:	gap of unknown length
9190:	contig of 1838 bp in length
9290:	gap of unknown length
11018:	contig of 1728 bp in length
11118:	gap of unknown length
13028:	contig of 1910 bp in length
13128:	gap of unknown length
14658:	contig of 1530 bp in length
14758:	gap of unknown length
16766:	contig of 2008 bp in length
16767:	gap of unknown length
16867:	contig of 2528 bp in length
19394:	gap of unknown length
19395:	gap of unknown length
22382:	contig of 2888 bp in length
22383:	gap of unknown length
22483:	contig of 4188 bp in length
26670:	gap of unknown length
26671:	contig of 3129 bp in length
29899:	gap of unknown length
29999:	gap of unknown length
32393:	contig of 2394 bp in length
32493:	gap of unknown length
35161:	contig of 2668 bp in length
35261:	gap of unknown length
35262:	contig of 3992 bp in length
39254:	gap of unknown length
42753:	contig of 3400 bp in length
42853:	gap of unknown length
45570:	contig of 2717 bp in length
45670:	gap of unknown length
50008:	contig of 4338 bp in length
50009:	gap of unknown length
50109:	contig of 5764 bp in length
55872:	gap of unknown length
55972:	gap of unknown length
60621:	contig of 4635 bp in length
60727:	gap of unknown length
64794:	contig of 4067 bp in length
64894:	gap of unknown length
69682:	contig of 4788 bp in length
69782:	gap of unknown length
74782:	contig of 5000 bp in length
74783:	gap of unknown length
80238:	contig of 5356 bp in length
80239:	gap of unknown length
86095:	contig of 5757 bp in length
86195:	gap of unknown length
93489:	contig of 7294 bp in length
93589:	gap of unknown length
99512:	contig of 5923 bp in length
93590:	

SOURCE	common honeybee.
ORGANISM	Mitochondrion Apis mellifera ligustica Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apoecrita; Aculeata; Apoidea; Apidae; Apis. 1 (bases 1 to 16343) Crozier.R.H. and Crozier.Y.C. The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization Genetics 133 (1), 97-117 (1993) 93114603
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
FEATURES	Location/Qualifiers 1..16343 /organism="Apis mellifera ligustica" /organelle="mitochondrion" /sub_species="ligustica" /db_xref="taxon:7469" /dev_stage="adult worker" /sex="female" /tissue_type="thorax" 1..66 /note="codon recognized: GAA" /product="tRNA-Glu" /anticodon=(pos:31..33,aa:Glu) 116..178 /note="codon recognized: AAU; anticodon unspecified" /product="tRNA-OTHER" 221..286 /note="codon recognized: AUG" /product="tRNA-Met" /anticodon=(pos:252..254,aa:Met) 296..358 /note="codon recognized: CAA" /product="tRNA-Gln" /anticodon=(pos:326..328,aa:Gln) 360..429 /note="codon recognized: GCA" /product="tRNA-Ala" /anticodon=(pos:393..395,aa:Ala) 433..501 /note="codon recognized: AAU; anticodon unspecified" /product="tRNA-OTHER" 502..1503 /codon_start=1 /transl_table=5 /product="NADH dehydrogenase subunit 2" /protein_id="AAB96798.1" /db_xref="GI:829004" /translation="MEFMFKYHWFYFTITFVLMMNSNTPIQWLMFEFGTIIS LINIKSTPSLIYSVSIVSIIFLFIIVLVSSISFTKDTDFNMVQMPLKIG TFPFHFMYTSPYEMMNKKOIFLMSLTIKPIIYMVMYSMTKINSWTLYFLITSLSYF YANKFTLAKLKACSTIFNSFYFIFILELNKNFIAMILYSFNYPILLISFLKNFIQ NEFMFYNKYQWFTLTMFNYSMPYFLSKNLNLFIMMYSVKAYNNIFFLLMISS MLMWNIILKRVLKMFNYKNNFIDDKNKTYHSYFALTULSNFFITLNF complement(1503..1571) /note="codon recognized: UGC" /product="tRNA-Cys" /anticodon=(pos:1536..1538,aa:Cys) complement(1592..1659) /note="codon recognized: UUU" /product="tRNA-Tyr" /anticodon=(pos:1621..1623,aa:Tyr) 1732..1793 /note="codon recognized: AAU; anticodon unspecified" /product="tRNA-OTHER" 1794..3359 /codon_start=1 /transl_table=5 /product="cytochrome c oxidase subunit 1" /protein_id="AAB96799.1" /db_xref="GI:552440" /translation="MKMKWMSTNHKNIGILIYIILAWSGLSGSMSLIIIRMEISSPGS WTSNDIQNTIVTSHAFLTFMFPFLGGGNWLPIPLMGSPDMAFRMNNISFWL
source	
tRNA	
tRNA	
tRNA	
tRNA	
tRNA	
CDS	
tRNA	
tRNA	
tRNA	
CDS	

LPPSLMILLSNLFPSPGTGWTVVPPLSAYLYHSSPSVDFAIFALSHMSGISIMGSS	
NLMVTIMMKNFSDMDQISLPFSWFITAILLIMSPLAGKAITMLFCNLSMIAMI	
DPWGDDIPPLYQHLEFWFGEPEVIILILPGFLISHIVMNSGKKELFCNLSMIAMI	
GICGLFIVWAHHMTVGLDVTDRATVTSATMIIAVPTGKVFWSLATVHGSKLKI	
SILWSLGFTMETTGCTGIMLSNSSIDILILDHTYVVVGHFHYVLSMGAVAFIISFI	
HNPYLITGLLNLIKWKIOFIMMFIGNVLTFFPHFGLGMSRPRTSDPYSDSYCNS	
ISMGSMSLSNSMIFLIFILLESLSKLRLPKFNOSSEWLNFLPPLPDHSHLEITPL	
INKNLKSILIKF"	
3355..3424	tRNA
/note="codon recognized: AAU; anticodon unspecified"	
/product="tRNA-OTHER"	
3618..4295	CDS
/codon_start=1	
/transl_table=5	
/product="cytochrome c oxidase subunit 2"	
/protein_id="AAB96800.1"	
/db_xref="GI:829005"	
/translation="MSTWFMFQESNSYYADNLISFHNVMMIIMIISTLTVTYIILD	
LFMKFSNFLKKNHNIEIWIITIPIILLIICFPCLKILYLIDEIVNPFSSIGH	
QMTWSEYEPENIEFDYMLANNLQFLLEDNRMVIPMKRILRLITTSTDVLIHS	
WVPSLGIRKVDAPVRINOLNISKRPIFFGQCSCICGMNHSFPMIESTSFQYFL	
NNVNKOI"	
4294..4362	tRNA
/note="codon recognized: GAC"	
/product="tRNA-Asp"	
/anticodon=(pos:4325..4327,aa:Asp)	
4370..4438	tRNA
/note="codon recognized: AAU; anticodon unspecified"	
/product="tRNA-OTHER"	
4444..4602	CDS
/codon_start=1	
/transl_table=5	
/product="ATPase subunit 8"	
/protein_id="AAB96801.1"	
/db_xref="GI:829006"	
/translation="MPQMPMKWFLIFYFYLLIFYFILINSMLTKTKINKETLKI	
LKKWNWF"	
4584..5264	CDS
/codon_start=1	
/transl_table=5	
/product="ATPase subunit 6"	
/protein_id="AAB96802.1"	
/db_xref="GI:552442"	
/translation="MKLILMNLFEMDPSTSNLSMNLMFMLPIIIFPSIFWLIQS	
RIMFKMTLMNFMYNFKVYQSNIIFISLMIYIMTNISLIPVYTLPTSL	
LLNMILSLTWFSFLIYLYNNYIMFLSHLVPVFLMNMVIELISLIIRPTL	
SIRLSANLSCHLILLLGIFISNIFISILPINLIONMLLTLEIFMSMQSYVFSILL	
IIFYSSSN"	
5285..6064	CDS
/codon_start=1	
/transl_table=5	
/product="cytochrome oxidase subunit 3"	
/protein_id="AAB96803.1"	
/db_xref="GI:552443"	
/translation="MRKNFPFMVTNSPWPILFSFPMNTLISTVIYSSISMFML	
NFTNSILIMLWPRDIIRSTFOGHSMFTNFKFSMILFILSELMFTISFPTFFH	
SSISPNEINMTWPKNIKPNPEIPLLNSPLTVSSGTTVTLSHVYLINNLUKLS	
YLLMTLLGIYFTILOITYESSNFFCFENDSIYCSIEFMATGPHGLHLVLCGISIFLLIS	
YRMNTHFSNMHNNFELAIIWHFVDVTLWLFDTIYILLI"	
6119..6184	tRNA
/note="codon recognized: GGA"	
/product="tRNA-Gly"	
/anticodon=(pos:6149..6151,aa:Gly)	
6185..6538	CDS
/codon_start=1	
/transl_table=5	
/product="NADH dehydrogenase subunit 3"	
/protein_id="AAB96804.1"	
/db_xref="GI:552444"	
/translation="MKFIFMYFIFIILISSILLLNKFISIYKKKDYEKSPCECFGN	
PITKANLPSPFLPFLMTMFLIFDVETIILFIPLIYFKSSSTMIYLSMISIFILLIT	
TLLEWNNYLNWLF"	
complement(6572..6638)	tRNA

```

/note="codon recognized: AAG"
/product="tRNA-Arg"
/anticodon={pos:6602..6604,aa:Arg}
6734..6802
/note="codon recognized: AAC"
/product="tRNA-Asn"
/anticodon={pos:6765..6767,aa:Asn}
complement(6810..6878)
/note="codon recognized: AUU; anticodon unspecified"
/product="tRNA-OTHER"
complement(6892..8556)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 5"
/protein_id="AAB96805.1"
/db_xref="GI:829007"
/translation="MIKMWCGILLPEFSFLMMLSLYLLYLKNEFFEPNWIYTNMS
RNFNLDYKSLMIFLVSMIFSMIIYSIMOLSELKMDRFILYMLFIDISIMML
ILSPNMLSIILGWDGLISYCLVIYVMKMSFTSGMVTILLNLGLDGLLLIMGLMT
YGSWNLSFKYKNEFMIIYILMAFTKSAQIPFTWLPMAKMAPTPVSLVHSSLTIVT
AGIYLLIRYVNLDPNKNYIMLIASLMLFAGLVANFELDLKVVAVSTLSQLGPM
SMLSIGSELVFLHLFIHAMFKSLMFCMCGSYMHVYSNODIRMYGYGYIYPMKSMI
LIFSILSCGPFVLGYYSKDLIIEMFFFSKMIYFSMNLIIIGTIFTVYSRSMILVL
TSKFLMNVITYSKEDIKCISMMMMIIFSLIYSKLIENLNMFNLLGIMLLIYKLMVF
KNIMVGLMGFNFYKLILNNKIKGYRMSFPMNLIYKIIYKKIIMMFTVEYIEKS
IIEILSSKFMSTLNIYELKISNMINIYTLIYLIYLLIYVINF"
complement(8557..8624)
/note="codon recognized: CCA"
/product="tRNA-His"
/anticodon={pos:8589..8591,aa:His}
complement(8644..9987)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/protein_id="AAB96806.1"
/db_xref="GI:552446"
/translation="MYLLLLMLMLMMSMIYLFMLPMKNMKNLNLIGNLIINIIL
LNLFLNLMNDIYIFCNLFNNYSYGLIMLFIWIFGLIFISLNNNSLNCFLPMNLLM
ISLLEFLNMLLLELYFEGLLIFLYLVKMGYSENRWSLGFYLMFTMFLSLPML
YIYIYILDISLNLMEMLNLNLMFLVLLMSFLVKIPIYLFHGLLKAHVEAP
YGSMLIASLMLGSGMLRMIIYKNEFILOKILVINSFGVLLISLMLCSOFDM
KSIILAISSIVHGLMHSMTFLKISLGGVLMISHGLSSGCLFELVNVISQNSR
LMFINKMGNSLMLWMLCNSGSPVSLNISEVMLLIGMISLWKFMLILMM
YCLFSFIYSIYLFMFINGKFIEMFKIRKINGILVEYVLLHWHIPLNLMFLKLYFI"
complement(9991..10254)
/codon_start=1
/transl_table=5

Query Match 3.1%; Score 72; DB 3; Length 16343;
Best Local Similarity 44.8%; Pred. No. 0.008;
Matches 361; Conservative 0; Mismatches 440; Indels 5; Gaps 2;

QY 1317 aaaaaattcaccacaataagagagacactacatctctcttattataataaaat 1376
|| |||| || || || || || || || || || || || || || || || || || ||
Db 15458 AATCAATTATAACAATTACTAAAAAATACAACTAGTATTATTAAATTTATAAT 15517
|| || || || || || || || || || || || || || || || || || || || ||
QY 1377 gtaagaaaaatagatataaaataacacatatcttctgataatttactactaaaactat 1436
|| || || || || || || || || || || || || || || || || || || || ||
Db 15518 AAATAATATAAATAATTTTAAATATATATATATATATATATATATATATATATAT 15577
|| || || || || || || || || || || || || || || || || || || || ||
QY 1437 ttctagtacttgtaatacatgctgagattttacagtaataaagaacagaggtagccc 1496
|| || || || || || || || || || || || || || || || || || || || ||
Db 15578 TTATTTATTCATAGTATTTAATATATAATTTTATTTAGTATAAAATTTTAAATATAA 15637
|| || || || || || || || || || || || || || || || || || || || ||
QY 1497 aaacaaaagtataatgtggagggtgtagctcttctcggtgcacaaaataaaccccaa 1556
|| || || || || || || || || || || || || || || || || || || || ||
Db 15638 AATTAACATTTTATATATAAATAATTTTTTTTATTAATTAATTTAAATATAA 15697
|| || || || || || || || || || || || || || || || || || || || ||
QY 1557 acttgtagatattgtgcgactgcctgcctgcctacattgaaatgaatgctctttat 1616
|| || || || || || || || || || || || || || || || || || || || ||
Db 15698 AATAATAAATGAGTTTTTTTTTTTGTATATATTTTATTATTATAAATAAATA 15757
|| || || || || || || || || || || || || || || || || || || || ||
QY 1617 aacgttgtctatgcgctattaccocataggtcactagaatgggacaatgaaat 1676
```

```

Db 15758 TATATATATATATAATATATATATATATATATATATATATATATATATATATAT 15817
|| || || || || || || || || || || || || || || || || || || || ||
QY 1677 atatctgcatgctggtggattcaatttaattgtagtcgtaaatggtagacatactca 1736
|| || || || || || || || || || || || || || || || || || || || ||
Db 15818 TATCTCTAATATATATTAATTTTAAATAAATATATATATATATATATATATATATAT 15877
|| || || || || || || || || || || || || || || || || || || || ||
QY 1737 tgcacacaaattatatactcactcgtgcaactcactcgtggtctctctcccatga 1796
|| || || || || || || || || || || || || || || || || || || || ||
Db 15878 TTAAATAATATATATATAATATATATATATATATATATATATATATATATATATAT 15937
|| || || || || || || || || || || || || || || || || || || || ||
QY 1797 attcacattgctaaagaataattaccaccttaaaatggtttatccctcgcacacattcca 1856
|| || || || || || || || || || || || || || || || || || || || ||
Db 15938 ATTAATTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 15997
|| || || || || || || || || || || || || || || || || || || || ||
QY 1857 tcaattattaaacattttaccattggaacacacacacacacacacacacacacacattttg 1916
|| || || || || || || || || || || || || || || || || || || || ||
Db 15998 TTTAAAAAATTCATATAACATATATATGATAATAAGCATAAT-AATTAATTTTTTAATA 16056
|| || || || || || || || || || || || || || || || || || || || ||
QY 1917 catttcaaaaactaaacacacacacacacacacacacacacacacacacacacatttca 1976
|| || || || || || || || || || || || || || || || || || || || ||
Db 16057 TTTTCTCTAATTAATAATTTTAAATAAAACTTAATAAAATAAAATAAAATTTTAA 16116
|| || || || || || || || || || || || || || || || || || || || ||
QY 1977 aaatatcctagtcttcaaccactcaataattcacacacacacacacacacacacacacac 2036
|| || || || || || || || || || || || || || || || || || || || ||
Db 16117 AAATAA---AACTTTAAAAATAAAAAATTTAAAAATAAAAACTTTTAAAAATAAAAA 16172
|| || || || || || || || || || || || || || || || || || || || ||
QY 2037 cacacactcagaacacttgattcaataataataaagaacacacacacacacacacacac 2096
|| || || || || || || || || || || || || || || || || || || || ||
Db 16173 AAAAACTTTTAAAAATAAAAACTTTTAAAAATAAAAAATTTAAAAATAAAAA 16232
|| || || || || || || || || || || || || || || || || || || || ||
QY 2097 atatacaccatatatgttatgataaa 2122
|| || || || || || || || || || || || || || || || || || || || ||
Db 16233 TTTTAAAAATAAAAAATAAAAA 16258
|| || || || || || || || || || || || || || || || || || || || ||

RESULT 8
LOCUS AC006970 136098 bp DNA PRI 27-SEP-2000
DEFINITION Homo sapiens clone RP4-725G10, complete sequence.
ACCESSION AC006970
VERSION AC006970.6 GI:10312289
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 136098)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 136098)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 136098)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 27, 2000 this sequence version replaced gi:9838025.
Center project name: HJ00725G10.
FEATURES
source
1..136098
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP4-725G10"
BASE COUNT 39294 a 29779 c 30262 g 37763 t
ORIGIN
```

[illegible]

Qy	2269	ccttttctaataatataactatcctcttggaacacctcttttgtatct	2315
Db	114030	TATTTTATATAATTATTTATGCTAGTATATATATATATATATATATAT	114076
RESULT 9			
AC006279	245802 bp DNA HPG	12-AUG-2000	
LOCUS	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.		
DEFINITION	AC006279		
ACCESSION	AC006279.8 GI:9797723		
VERSION	HG; HTGS-PHASEL		
KEYWORDS	malaria parasite P. falciparum.		
SOURCE	Plasmodium falciparum Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
ORGANISM	1 (bases 1 to 245802)		
REFERENCE	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J.J., Tamaki,T., Kurdi,O.B., Conway,A.B. and Davis,R.W.		
AUTHORS	Kurdi,O.B., Conway,A.B. and Davis,R.W.		
TITLE	Plasmodium falciparum 3D7 chromosome 12 Unpublished		
JOURNAL	2 (bases 1 to 245802)		
REFERENCE	Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.		
AUTHORS	Direct Submission		
TITLE	Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA		
JOURNAL	94304, USA		
COMMENT	On Aug 12, 2000 this sequence version replaced gi:8810451. * NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved		
FEATURES	* 1 147568: contig of 147568 bp in length * 147569 147768: gap of unknown length * 147769 217470: contig of 69702 bp in length * 217471 217670: gap of unknown length * 217671 219372: contig of 1702 bp in length * 219373 219572: gap of unknown length * 219573 221333: contig of 1761 bp in length * 221334 221533: gap of unknown length * 221534 245802: contig of 24269 bp in length. Location/Qualifiers 1..245802 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /chromosome="12" /clone="PFYAC614" /Clone="3D7"		
BASE COUNT	99718 a 25978 c 21493 g 97813 t	800 others	
ORIGIN			
Query Match	3.0%; Score 68.6; DB 2; Length 245802;		
Best Local Similarity	45.5%; Pred. No. 0.023;		
Matches	401; Conservative 0; Mismatches 469; Indels 11; Gaps 4;		
Qy	1249	tctgataccattaagatcaacttcgtgatggatcatcattcacatatcataaacatttac	1308
Db	111408	TATTATGTATTATATATATATATATATGTCGAAGTGTAATAAATTTATATATATATGATATT	111467
Qy	1309	gtaagaataaaaaattcccacaacaagttagagagagacactacatctctcttatatt	1368
Db	111468	TATTATATATAAAACTGTTTGATATAAAAATGGAAAAAANAATTTATGAAGATGAAAAATT	111527
Qy	1369	aataaaatlgtaaaaga-----aaaaatagataaaaagaatcacatatcttttgataaatatttatt	1425
Db	111528	TCAAAATTTTATAAACAGAAANAATAGCATATTTACATTTATATATATATATATATATATAAT	111587
Qy	1426	actaaaactattttcttagtacttgttaatcatgctcgaggatgttttaccagtaaataaaa	1485

clone_end:T7 vector_side:right" BASE COUNT 40346 a 38964 c 40180 g 40496 t 802 others ORIGIN	Query Match 2.9%; Score 68; DB 2; Length 160788; Best Local Similarity 43.5%; Pred. No. 0.03; Matches 363; Conservative 0; Mismatches 465; Indels 6; Gaps 1;	AL593856.6 GI:15041959 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 156133) Mclay, K. Direct Submission Submitted (28-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerquest@sanger.ac.uk On Jul 30, 2001 this sequence version replaced gi:15028764. ----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: BA8J9 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 153103 bases at least Q40 Consensus quality: 153756 bases at least Q30 Consensus quality: 154302 bases at least Q20 Insert size: 155533; sum-of-contigs Insert size: 163653; 1.8% error; agarose-fp Quality coverage: 6.73x in Q20 bases; sum-of-contigs Quality coverage: 6.48x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 7 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 3197: contig of 3197 bp in length * 3198 3297: gap of 100 bp * 3298 24805: contig of 21508 bp in length * 24806 24905: gap of 100 bp * 24906 27453: contig of 2548 bp in length * 27454 27553: gap of 100 bp * 27554 64419: contig of 36866 bp in length * 64420 64519: gap of 100 bp * 64520 84883: contig of 20364 bp in length * 84884 84983: gap of 100 bp * 84984 133876: contig of 48893 bp in length * 133877 133976: gap of 100 bp * 133977 156133: contig of 22157 bp in length.
		Location/Qualifiers 1..156133 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /clone="RP11-8J9" /clone_lib="RPC1-11.1" 1..3197 /note="assembly_fragment:00456 fragment_chain:1" 3298..24805 /note="assembly_fragment:00590 fragment_chain:1" 24906..27453 /note="assembly_fragment:01416 fragment_chain:1" 27554..64419 /note="assembly_fragment:03269 fragment_chain:1" 64520..84883
		FEATURES source misc_feature misc_feature misc_feature misc_feature misc_feature
clone_end:T7 vector_side:right" BASE COUNT 40346 a 38964 c 40180 g 40496 t 802 others ORIGIN	Query Match 2.9%; Score 68; DB 2; Length 160788; Best Local Similarity 43.5%; Pred. No. 0.03; Matches 363; Conservative 0; Mismatches 465; Indels 6; Gaps 1;	AL593856/c LOCUS DEFINITION Homo sapiens chromosome 1 clone RP11-8J9, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces. ACCESSION AL593856

Qy	2087	tatctaaacaataacacocatatgttattgatataataatgcatgcagcaatacacttaatt	2146
Dd	27002	ATATATAAAAAATACAAAATTTATATATATAAAATATAAATATAAAAAATAAATAAAT	26943
Qy	2147	tggtaaagcattaaagcgagacaactctattaaaccoggttaattccaaca	2195
Dd	26942	ATATATAATTTATATAATAAATAATATATATATATACATATATAAATACATA	26894

RESULT 12
AC072048/c

LOCUS AC072048 208684 bp DNA HTG 07-JUN-2000
DEFINITION Mus musculus chromosome 6 clone RP23-189E15 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 24 unordered pieces.

ACCESSION AC072048
VERSION 1 GI-8313202
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mu

REFERENCE 1 (bases 1 to 208684)
AUTHORS Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Lin,S.-O., Maduro,Q.L., Maduro,V.B.,
Masterian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.

TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208684)
AUTHORS Green,E.D.
DIRECT SUBMISSION
TITLE Submitted (07-JUN-2000) NIH Intramural Sequencing Center, 8717
JOURNAL Grovemont Circle, Gaithersburg, MD 20877, USA

COMMENT ----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgri.nih.gov
----- Project Information
Center project name: wp
Center clone name: 189E15
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195465 bases at least Q40
Consensus quality: 197294 bases at least Q30
Consensus quality: 198446 bases at least Q20
Insert size: 209000; agarose-fp
Insert size: 197000; pulse-field-gel
Insert size: 206384; sum-of-contigs
Quality coverage: 6.43x in Q20 bases; agarose-fp
Quality coverage: 6.82x in Q20 bases; pulse-field-gel
Quality coverage: 6.51x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	2209:	contig of 2209 bp in length
*	2210	2309:	gap of unknown length
*	2310	4736:	contig of 2427 bp in length
*	4737	7838:	gap of unknown length
*	4837	7023:	contig of 2187 bp in length
*	7024	7123:	gap of unknown length
*	7124	9503:	contig of 2382 bp in length

```

* 9506 9605: gap of unknown length
* 9606 14706: contig of 5101 bp in length
* 14707 14806: gap of unknown length
* 14807 18659: contig of 3853 bp in length
* 18660 18759: gap of unknown length
* 18760 23424: contig of 6665 bp in length
* 23425 25525: gap of unknown length
* 25525 33399: contig of 7775 bp in length
* 33399 33949: gap of unknown length
* 33949 39450: contig of 6050 bp in length
* 39450 39549: gap of unknown length
* 39549 48515: contig of 8966 bp in length
* 48515 50233: gap of unknown length
* 50233 55123: contig of 6408 bp in length
* 55123 64442: gap of unknown length
* 64442 73027: contig of 9319 bp in length
* 73027 73127: gap of unknown length
* 73127 82147: contig of 9020 bp in length
* 82147 82247: gap of unknown length
* 82247 93854: contig of 11607 bp in length
* 93854 93955: gap of unknown length
* 93955 103237: contig of 9282 bp in length
* 103237 117502: gap of unknown length
* 117502 117602: contig of 14166 bp in length
* 117602 130162: gap of unknown length
* 130162 130262: contig of 12560 bp in length
* 130262 140708: gap of unknown length
* 140708 151652: contig of 10446 bp in length
* 151652 151752: gap of unknown length
* 151752 163823: contig of 10844 bp in length
* 163823 163923: gap of unknown length
* 163923 178830: contig of 12071 bp in length
* 178830 178930: contig of 14807 bp in length
* 178930 193909: gap of unknown length
* 193909 194009: contig of 15079 bp in length
* 194009 208684: gap of unknown length
* 208684 194010: contig of 14675 bp in length.
FEATURES             Location/Qualifiers
     source            1..208684
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /clone="RP23-189E15"
                     /clone_lib="RPCI mouse BAC library 23"
misc_feature          1..2209
misc_feature          2310..4736
misc_feature          /note="assembly_fragment"
misc_feature          4837..7023
misc_feature          /note="assembly_fragment"
misc_feature          7124..9505
misc_feature          /note="assembly_fragment"
misc_feature          9606..14706
misc_feature          /note="assembly_fragment"
misc_feature          14807..18659
misc_feature          /note="assembly_fragment"
misc_feature          18760..23424
misc_feature          /note="assembly_fragment"
misc_feature          23425..33399
misc_feature          /note="assembly_fragment"
misc_feature          33400..39449
misc_feature          /note="assembly_fragment"
misc_feature          39450..39549
misc_feature          /note="assembly_fragment"
misc_feature          39550..48515
misc_feature          /note="assembly_fragment"
misc_feature          48516..50233
misc_feature          /note="assembly_fragment"
misc_feature          50234..55123
misc_feature          /note="assembly_fragment"
misc_feature          55124..64442
misc_feature          /note="assembly_fragment"
misc_feature          64443..73027
misc_feature          /note="assembly_fragment"

```

```

misc_feature          /note="assembly_fragment"
misc_feature          73128..82147
misc_feature          /note="assembly_fragment"
misc_feature          82248..93854
misc_feature          /note="assembly_fragment"
misc_feature          93955..103236
misc_feature          /note="assembly_fragment"
misc_feature          103337..117502
misc_feature          /note="assembly_fragment"
misc_feature          117603..130162
misc_feature          /note="assembly_fragment"
misc_feature          130263..140708
misc_feature          /note="assembly_fragment"
misc_feature          140809..151652
misc_feature          /note="assembly_fragment"
misc_feature          151753..163823
misc_feature          /note="assembly_fragment"
misc_feature          163924..178730
misc_feature          /note="assembly_fragment"
misc_feature          clone_end:T7
misc_feature          vector_side:right
misc_feature          178831..193909
misc_feature          /note="assembly_fragment"
misc_feature          clone_end:SP6
misc_feature          vector_side:right
misc_feature          194010..208684
misc_feature          /note="assembly_fragment"

BASE COUNT  54390 a 48824 c 48605 g 54542 t 2323 others
ORIGIN

Query Match      2.9%; Score 67.6; DB 2; Length 208684;
Best Local Similarity 46.5%; Pred.No.0.034;
Matches 240; Conservative 0; Mismatches 275; Indels 1; Gaps 1;

QY 1667 aatttaataatctgtcgtggtgggttcatttaattgtatcgtaaatgtag 1726
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208567 ATTATAATAAATAATAATAATAATAATAATAATAATAATAATAATAATA 208508

QY 1727 gacatactcatgcacaaattatcatcatcactggcgaactcactggtcgaatggtttct 1786
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208507 TTTATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 208448

QY 1787 ctcccatgaattccattgctcaagaataattaccaccttaaaatgtttatccctgcac 1846
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208447 ATATAAATTTATTAATAATAATAATAATAATAATAATAATAATAATAATA 208388

QY 1847 acattcacatcaattattataaacattttaccattggaaacacacatacatatcaatca 1906
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208387 TAAATTTAATTTTATTATAATATATATAATAATAATAATAATAATAATAATA 208328

QY 1907 attattttgcatctttcaaaactaaacccaacaaacttagaataatttggtaattatagc 1966
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208327 AAAATATATATAATTTTAAAAATAAAAAATATATATAATAATAATAATAATA 208268

QY 1967 acaattttcaaaaatccatcctagctctcaaccactcaattcacaattcccaat-ccc 2025
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208267 ATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 208208

QY 2026 ttgcaaaacatcacacacctctagaaactttgtatttaataacttaataaaagcaataatag 2085
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208207 TTTTAATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 208148

QY 2086 atatctaaacaatatccaccatatgtgtatgatataataatgatgcgcgaatacacaataat 2145
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208147 ATAATATAAAATTTATATAATAATAATAATAATAATAATAATAATAATAATA 208088

QY 2146 ttggtaaacgattaaagcgagacaactctattaaca 2181
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208087 TAAATAAAATAATAAAATTTATATAATAATAATAATAATAATAATAATAATA 208052

RESULT 13

```

```
AC005506      205429 bp      DNA      12-AUG-2000
LOCUS          Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
DEFINITION    PROGRESS ***, 4 unordered pieces.
ACCESSION     AC005506
VERSION       AC005506.8 GI:9797717
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE        malaria parasite P. falciparum.
ORGANISM      Plasmodium falciparum
REFERENCE     1 (bases 1 to 205429)
AUTHORS       Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
              Kurdi,O.B., Conway,A.B. and Davis,R.W.
TITLE         Plasmodium falciparum 3D7 chromosome 12
REFERENCE     2 (bases 1 to 205429)
AUTHORS       Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE         Direct Submission
JOURNAL       Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
              Center, Stanford University, 855 California Avenue, Palo Alto, CA
              94304, USA
COMMENT       On Aug 12, 2000 this sequence version replaced gi:8810449.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 4 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 132269: contig of 132269 bp in length
              * 132270 132469: gap of unknown length
              * 132470 192227: contig of 59758 bp in length
              * 192228 192427: gap of unknown length
              * 192428 203864: contig of 11437 bp in length
              * 203865 204064: gap of unknown length
              * 204065 205429: contig of 1365 bp in length.
              *
              * Location/Qualifiers
              *   .205429
              *   /organism="Plasmodium falciparum"
              *   /db_xref="taxon:5833"
              *   /chromosome="12"
              *   /clone="PFYAC357"
              *   /clone="3D7"

BASE COUNT    83932 a 17937 c 19766 g 83194 t 600 others
ORIGIN

Query Match    2.9%; Score 66.6; DB 2; Length 205429;
Best Local Similarity 43.4%; Pred. No. 0.049;
Matches 356; Conservative 0; Mismatches 464; Indels 1; Gaps 1;

QY 1350 acatctcttattataataaagttaaaagaaataatagataaaagtaaacacata 1409
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83502 AATCAGATATGTCATTAAATAATAATAATATATATATA-TATATATATTATGTT 83560
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1410 ttttgataaattattactaaaactatttcttagtactgttaactcatgctgaggattt 1469
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83561 TTTTATTATTATTCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 83620
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1470 tacagtaataagaacagggtagcccaacaaagtgataattgtggaggggtgatct 1529
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83621 CTATTTAGGAAATAAATATATATAAGAAATAAATAAATAAATAAATAAATAAATAA 83680
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1530 ttgtcggtcgcaaaatgaaccccaactgtgatgtgtgctgactgctcgctgcta 1589
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83681 ATAAATAGATAATAAAGAACATCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 83740
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1590 cattgaaataatgaattcttttataacgtttgtctatgctgcgtattaccctatggtc 1649
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83741 AATTAATAATTAAGAAATATAAAATTTATATACATATATATATATATATATATATA 83800
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1650 actagaatgggacaatgaatttaataatatatctgtcatgtgtgtgggttcatttaatt 1709
```

```
Db 83801 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 83860
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1710 tgcataatgtaggacatactactgcacacattatcatcactggtgcaatcac 1769
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83861 ACAGTGATAAATATTAATATATATATACATATATATATATATATATATATATATATATA 83920
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1770 tggccaatgtgtttctcttcccatgaattccacattgctgtaagaaataaccaccttaaa 1829
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83921 TTATAATTAATAATAGCATACTATAATTAATTTCTTTTCTTTTCTTTTCTTTTCTTTT 83980
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1830 atgtttatccctgcacacatttccacatcaatttattaaacattttccaccattggaaaac 1889
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83981 ATATTTATGCATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 84040
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1890 acatacatattcaatcaattatttttgcattttcacaacataacacaaacaaacttagaa 1949
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84041 TTCAAAATTTTATTTATATATGCTATTCATAAAAGTATATATGTTGAAATACACACAAA 84100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1950 tatttgaattatagcacattttcacaataatccctagcttccacacctcaaatc 2009
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84101 AAAAAAGGAAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 84160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2010 acaatttccaaatccctgcacacatacacacacctctagaaactttgtattaataactaa 2069
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84161 ATAAAAATGAAAAATTTTAATAAATTCATGATATATATATATATATATATATATATACAAT 84220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2070 taaaagcaataatgatatactcaacaataatccacataatgtttatgataatgatg 2129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84221 ATTCAATAATATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 84280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2130 cagcaatacacacttaatttggtaagcattaaagcgagacaa 2170
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84281 TTCTTATAAGCAATTTTGAGTAAATATATATAAATAAGGAAA 84321
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
PFMAL4P4      224448 bp      DNA      HTG
LOCUS          Plasmodium falciparum chromosome 4 strain 3D7, *** SEQUENCING IN
DEFINITION    PROGRESS ***, in unordered pieces.
ACCESSION     AL035477
VERSION       AL035477.5 GI:5731932
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE        malaria parasite P. falciparum.
ORGANISM      Plasmodium falciparum
REFERENCE     1 (bases 1 to 224448)
AUTHORS       Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
              Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
              and Barrell,B.
TITLE         Direct Submission
JOURNAL       Submitted (24-FEB-1999) P.falciparum Genome Sequencing Consortium,
              The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA, UK
COMMENT       On Aug 12, 1999 this sequence version replaced gi:5531400.
              For more information about this sequence or the Malaria Project,
              see http://www.sanger.ac.uk/Projects/P.falciparum. IMPORTANT: This
              sequence is unfinished and does not necessarily represent the
              correct sequence. Work on the sequence is in progress and the
              release of this data is based on the understanding that the
              sequence may change as work continues. The sequence may be
              contaminated with foreign sequence from E.coli, yeast, vector,
              phage etc.
              Order of segments is not known; 800 n's separate segments.
              * NOTE: This is a 'working draft' sequence.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * Location/Qualifiers
              *   1..224448
              *   /organism="Plasmodium falciparum"
              *   /strain="3D7"
              *   /db_xref="taxon:5833"
```

BASE COUNT	80792	a	23258	c	21957	g	75240	t	23201	others
ORIGIN	/chromosome="4"									

Query Match 2.9%; Score 66.6; DB 2; Length 224448;
Best Local Similarity 20.1%; Pred. NO. 0.049;
Matches 282; Conservative 0; Mismatches 1114; Indels 7;

Matches	282;	Conservative	0;	Mismatches	1114;	Indels	7;	Gaps	2;
Qy	36	aaataataaaactcctaaagaatgattttattcttattcttctcctaataaactttccctatt	95						
Db	72021	AAATTTTACCTTTCTATCTATATTTATCTACCATATATCTATGAAATAAATACATATATA	72080						
Qy	96	ccaaaacacatcaaaagttagtgatcctctttaaattctgataataataattgt	155						
Db	72081	CGATAGATATATATTTTTAGGTATTTTATCATATTTGAAATAGTATATTATAAATATTA	72140						
Qy	156	atattcaatatctcatacaattgttgtttatgaaatatctttagggaaaaaggactaaag	215						
Db	72141	ATTTTATTGTTATATAATATAGTTTTATAATATATTACAATATATAATGTATTATATAGT	72200						
Qy	216	ataaactccgcacaactcaaaagtcagaaacccctgttaactcttcagttgaaacgagaag	275						
Db	72201	ATATATGTACAACTACAAAATAAATAAATAAATAAT-ACAATAACAATAACATAGAGAA	72259						
Qy	276	gaagtggacaacacagaaaactaaagttcccccacttaactcttggttggtggaggact	335						
Db	72260	GGACTATTGTATTATAAATAAATAATACATTTTTTTATTCTATCTTTTATATACTTCA	72319						
Qy	336	tcctttcaaatctatactctaaaggaaaacacattagacaactctagatgggtgcacttagct	395						
Db	72320	TTTTATATAATTTATATTTTATGTATTCATGTACTTATATATCTCCCTTATTTTGT	72379						
Qy	396	catatatttttaagdaataacacccactccaagttttttgtttgtgtgtgcagta	455						
Db	72380	C-----TTTATAATAAAAAAGACATTAAGTGACAAATTTATATATGTATANNNNNN	72433						
Qy	456	gatgataagatggatcattctcgaagcccttatgcaagacataagaatccatactcc	515						
Db	72434	NN	72493						
Qy	516	accaagatgctttacatctcaaccaagttaatgaatttaactcttcgaaacaattatt	575						
Db	72494	NN	72553						
Qy	576	cctccaaaagaagtttatatgcacattctcctaatgtatttatataagaattgatacat	635						
Db	72554	NN	72613						
Qy	636	gttctgttatacaagattagaatttggtattctcatcctcaactcctacacttggtgaga	695						
Db	72614	NN	72673						
Qy	696	aatttcagctcaacctcagtaaatcagggtctctcctcctcaactcatacacttggtgag	755						
Db	72674	NN	72733						
Qy	756	tgaagaattatggacgtcaacctagcaatatgaatccctcctcgaagatcctcacacttact	815						
Db	72734	NN	72793						
Qy	816	gagtgaagaatttggtctctcgacctcaacaagatagatttgatgggtcatcacgagggga	875						
Db	72794	NN	72853						
Qy	876	agcattcacattgggtcaaaagattcaccccaacaagtgagagagacatcacatcaacc	935						
Db	72854	NN	72913						
Qy	936	aaaaccttaagtgatagggtgatagttctcttacttataaagtgtcaccctccactt	995						
Db	72914	NN	72973						

[illegible]

RESULT	15
--------	----

LOCUS	DEFINITION	SEQUENCE	UNORDERED	PIECES	HTG	03-JUL-2001
AC092304	Homo sapiens chromosome 19 clone LLNLFOS-20C5, WORKING DRAFT	36977 bp	DNA			

ACCESSION
AC092304

VERSION AC092304.1 GI:14589493
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN
SOURCE human.

SOURCE
ORGANISM

REFERENCE

1 (bases 1 to 36977)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS
DOE Joint Genome Institute
IN (bases 1 to 30977)

ACTIONS	DOE UTILIC G	Sequencing	TITLE

JOURNAL		Unpublished
TITLE		sequentially

REFERENCE 2 (bases 1 to 36977)

REFERENCE
 AUTHORS
 DOE Joint Genome Institute.
 (pages 1 to 2037)

TITLE Direct Submission

JOURNAL Submitted (03-JUL-2001) Product

Genome Institute, 2800 Mitchel

COMMENT
-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

1
2
3
4
5
6
7

Project Information

Center Project Name: 31883, FO

Center clone name: LLNL-FOS_20

[illegible]

Summary Statistics

Consensus quality: 311111 bases

Consensus quality: 33034 bases

Consensus quality: 338/1 bases
Estimated length: 42000

Estimated insert size: 42980;
Estimated insert size: 26477;

Estimated insert size: 36477;
Quality coverage: 55548 030

Quality coverage:	6.56	in Q20
Quality coverage:	7.73	in Q20

* NOTE: This is a 'working draught' quality coverage: 7:73 III Q20

* consists of 6 contigs. The t

convergence: the

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:04:45 ; Search time 1999.71 Seconds
(without alignments)
98.997 Million cell updates/sec

Title: US-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.om.*

20: em.or.*

21: em.ov.*

22: em.pat.*

23: em.ph.*

24: em.pl.*

25: em.ro.*

26: em.sts.*

27: em.sy.*

28: em.un.*

29: em.vi.*

30: em.htgo.hum.*

31: em.htgo.inv.*

32: em.htgo.rod.*

33: em.htg.hum.*

34: em.htg.inv.*

35: em.htg.rod.*

36: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	12	100.0	22	6	AX004331	Sequence
C 2	12	100.0	145	5	CHRMIL08	K03266 Chicken c-m
C 3	12	100.0	164	11	G44227	G44227 WIAF-3873-S
C 4	12	100.0	164	11	G59874	G59874 SHGC-130652
C 5	12	100.0	215	9	HSA006490	AJ006490 Homo sapi
C 6	12	100.0	220	9	APU09224	U09224 Alouatta pa
C 7	12	100.0	222	11	G48131	G48131 SMSS31 Mous
C 8	12	100.0	228	14	AU038679	U38679 Avian infec
C 9	12	100.0	231	14	AU038677	U38677 Avian infec
C 10	12	100.0	231	14	AU038678	U38678 Avian infec
C 11	12	100.0	231	14	AU038681	U38681 Avian infec
C 12	12	100.0	285	11	AU025328	U025328 Rattus no
C 13	12	100.0	366	14	HPCTH13P30	D28849 Hepatitis C
C 14	12	100.0	368	14	IBR243261	AJ243261 Infectiou
C 15	12	100.0	388	11	G30656	G30656 human STS S
C 16	12	100.0	454	11	G56069	G56069 SHGC-101332
C 17	12	100.0	463	11	HUMUT5361	L18645 Human chrom
C 18	12	100.0	475	11	G51167	G51167 SHGC-80775
C 19	12	100.0	679	6	A86412	A86412 Sequence 10
C 20	12	100.0	679	6	ARI55905	ARI55905 Sequence
C 21	12	100.0	679	6	E66430	E66430 Genome DNA
C 22	12	100.0	681	2	AC061540	AC061540 Giardia i
C 23	12	100.0	724	6	AX187114	AX187114 Sequence
C 24	12	100.0	725	5	AF231326	AF231326 Nyctea sc
C 25	12	100.0	725	5	AF231331	AF231331 Asio otus
C 26	12	100.0	740	2	AC038063	AC038063 Giardia i
C 27	12	100.0	776	2	AC015328	AC015328 Drosophil
C 28	12	100.0	802	8	AF363433	AF363433 Fusarium
C 29	12	100.0	803	8	AF363432	AF363432 Fusarium
C 30	12	100.0	919	9	AF069970	AF069970 Nasalis l
C 31	12	100.0	920	2	AC082280	AC082280 Giardia i
C 32	12	100.0	930	8	AF105062	AF105062 Phytolacc
C 33	12	100.0	935	9	BC008190	BC008190 Homo sapi
C 34	12	100.0	954	2	AC081006	AC081006 Giardia i
C 35	12	100.0	980	2	AC032864	AC032864 Giardia i
C 36	12	100.0	1000	2	AC062314	AC062314 Giardia i
C 37	12	100.0	1023	2	AC064712	AC064712 Giardia i
C 38	12	100.0	1080	2	AC062313	AC062313 Giardia i
C 39	12	100.0	1097	2	AC039640	AC039640 Giardia i
C 40	12	100.0	1138	6	AR083280	AR083280 Sequence
C 41	12	100.0	1138	6	AX099541	AX099541 Sequence
C 42	12	100.0	1188	6	AX011663	AX011663 Sequence
C 43	12	100.0	1307	3	AF216214	AF216214 Drosophil
C 44	12	100.0	1308	3	AF318611	AF318611 Caenorhab
C 45	12	100.0	1323	8	AF238237	AF238237 Arabidops

ALIGNMENTS

RESULT 1	AX004331/c	22 bp	DNA	PAT	24-AUG-2000
LOCUS	AX004331	Sequence 83 from Patent WO9919492.			
DEFINITION	AX004331				
ACCESSION	AX004331.1	GI:9927813			
VERSION					
KEYWORDS		synthetic construct.			
SOURCE		synthetic construct.			
ORGANISM		artificial sequence.			
REFERENCE		1 (bases 1 to 22)			
AUTHORS		Betzner, A.S. and Doutriaux, M.P.			
TITLE		Methods for obtaining plant varieties			
JOURNAL		Patent: WO 9919492-A 83 22-APR-1999.			
FEATURES		BETZNER ANDREAS STEFAN (AU); DOUTRIAUX MARIE PASCALE (FR)			
source		Location/Qualifiers			
		1..22			
		/organism="synthetic construct"			
		/db_xref="taxon:32630"			
		/note="Reverse primer for PCR amplification of NGA249 SSLP"			

```

BASE COUNT      7 a      7 c      3 g      5 t
ORIGIN
marker in Arabidopsis thaliana subspecies"

Query Match      100.0%; Score 12; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 5.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
   |||
Db 21 GGATTTTACAGT 10

RESULT 2
LOCUS CHMIL08/c 145 bp DNA VRT 28-APR-1993
DEFINITION Chicken c-mil proto-oncogene, exon 8.
ACCESSION K03266
VERSION K03266.1 GI:212316
KEYWORDS c-myc proto-oncogene; mil oncogene; proto-oncogene.
SEGMENT 8 of 11
SOURCE Chicken DNA, clones lambda c-mil-4 and lambda c-mil-7.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Jansen.H.W. and Bister,K.
AUTHORS 1 (bases 1 to 145)
TITLE Nucleotide sequence analysis of the chicken gene c-mil, the
        progenitor of the retroviral oncogene v-mil
JOURNAL Virology 143, 359-367 (1985)
MEDLINE 86045899
COMMENT See segment 1.
FEATURES
    Location/Qualifiers
        1..145
            /organism="Gallus gallus"
            /db_xref="taxon:9031"
            <1..20
                /note="c-mil intron H"
            21..139
                /note="c-mil protein, (putative); putative"
            /number=8
            140..>145
                /note="c-mil intron I"
BASE COUNT      37 a      27 c      38 g      43 t
ORIGIN About 900 kb after segment 7.

Query Match      100.0%; Score 12; DB 5; Length 145;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
   |||
Db 85 GGATTTTACAGT 74

RESULT 3
LOCUS G44227/c 164 bp DNA STS 28-JAN-1999
DEFINITION WIAF-3873-STS Human Thudson SANGER Homo sapiens STS genomic,
        sequence tagged site.
ACCESSION G44227
VERSION G44227.1 GI:4193144
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164)
REFERENCE Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
AUTHORS Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,

Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,S., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
98248615
Synonyms: stsg28550
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: TTGTGATAGCTGGTGAGGTGC
Primer B: TGGTGCCTTATCCCAACATT
STS size: 164
PCR Profile:
    Presoak: 94 degrees C for 4.00 minutes
    Denaturation: 94 degrees C for 50.0 seconds
    Annealing: 58 degrees C for 1.50 minutes
    Polymerization: 72 degrees C for 1.00 minutes
    PCR Cycles: 30
    Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
    Template: 10 ng
    Primer: each 5 pM
    dNTPs: 4 nM
    Taq Polymerase: 0.5 U
    Total Vol: 20 uL

Buffer:
    Mg2+: 1.5 mM
    KCl: 50 mM
    Tris-HCl: 10 mM
    Gelatin: .001 %
    Location/Qualifiers
        1..164
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="68.70 cR from top of Chr19 linkage group"
            /clone_lib="Human Thudson SANGER"
            /note="human STS created from EST in the Sanger database"
        1..164
        primer_bind 1..21
        primer_bind 1..21
        BASE COUNT 43 a 25 c 36 g 59 t 1 others
        ORIGIN

Query Match      100.0%; Score 12; DB 11; Length 164;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
   |||
Db 88 GGATTTTACAGT 77

RESULT 4
LOCUS G59874/c 164 bp DNA STS 30-MAR-2000
DEFINITION SHGC-130652 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION G59874
VERSION G59874.1 GI:6125193
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164)
REFERENCE

```


Submitted (22-MAY-1998) Ugarte M., Universidad Autonoma de Madrid,

```

Db 81 GGATTTTACAGT 70

RESULT 7
G48131/c
LOCUS      222 bp      DNA      STS      09-APR-1999
DEFINITION Mouse Eggen Mus musculus STS genomic, sequence tagged site.
ACCESSION  G48131
VERSION     G48131.1  GI:4494739
KEYWORDS   STS.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 222)
AUTHORS   Desilva,U., Massa,H., Trask,B.J. and Green,E.D.
TITLE     Comparative Mapping of the Region of Human Chromosome 7 Deleted in
            William's Syndrome
JOURNAL   Unpublished (1999)
COMMENT   Synonyms: 92N10R
            Contact: Eric D. Green
            Genome Technology Branch
            National Human Genome Research Institute/NIH
            49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892
            Tel: 3014020201
            Fax: 3014024735
            Email: egreen@nhgri.nih.gov
            Primer A: TGCCTCTAGTCACAGGATTGGC
            Primer B: AGATGCTTGGTTGGTAGGGTG
            STS size: 222
            PCR Profile:
                Presoak: 0 degrees C for 0.00 minute(s)
                Denaturation: 92 degrees C for 0.17 minute(s)
                Annealing: 58 degrees C for 1.00 minute(s)
                Polymerization: 72 degrees C for 1.00 minute(s)
                PCR Cycles: 35
            Thermal Cycler: PerkinElmer 9600
            Protocol:
                Template: 30-100 ng
                Primer: each 1 uM
                dNTPs: each 200 uM
                Taq Polymerase: 0.05 units/ul
                Total Vol: 10 ul

Buffer:
MgCl2: 1.5 mM
KCl: 100 mM
Tris-HCl: 10 mM
NH4Cl: 5 mM
pH: 8.6

The sequence for this STS was derived from a single sequencing
read.

FEATURES             Location/Qualifiers
     source           1..222
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="Mouse EGgen"
     STS              1..222
     primer_bind      1..222
     primer_bind      complement(201..222)
BASE COUNT          64 a 61 c 42 g 53 t 2 others
ORIGIN
1 ggaatttcacagt 12
|||||
57 GGATTTTACAGT 46

Query Match          100.0%; Score 12; DB 11; Length 222;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatttcacagt 12
|||||
Db 57 GGATTTTACAGT 46

RESULT 9
AIU38677
LOCUS      231 bp      RNA      VRL      14-FEB-1997
DEFINITION Avian infectious bronchitis virus Al211 spike protein (S1) gene,
            hypervariable region 1, partial cds.
ACCESSION  U38677
VERSION     U38677.1  GI:1055301
KEYWORDS
SOURCE     Avian infectious bronchitis virus.
            Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
            Coronaviridae; Coronavirus.
ORGANISM   1 (sites)
            Wang,C.H. and Tsai,C.T.
            Genetic grouping for the isolates of avian infectious bronchitis
            virus in Taiwan
JOURNAL   Arch. Virol. 141 (9), 1677-1688 (1996)
MEDLINE   97049060
REFERENCE  2 (bases 1 to 231)
AUTHORS   Wang,C.
TITLE     Direct Submission
JOURNAL   Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
            Veterinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
            Location/Qualifiers
            1..228
            /organism="Avian infectious bronchitis virus"
            /strain="Al171"
            /db_xref="taxon:11120"
            1..228
            /gene="S1"
            <1..228
            /gene="S1"
            /note="hypervariable region 1; N-terminus"
            /codon_start=2
            /product="spike protein"
            /protein_id="AAB47437.1"
            /db_xref="GI:1055306"
            /translation="GWHLGGAYAVVIFFTNNAGASECTNGIISGGSGFNASSIAM
            TAPVGPQMWSKSFQTAHCNCFSDFTVETH"
BASE COUNT          54 a 40 c 54 g 80 t
ORIGIN
1 ggaatttcacagt 12
|||||
202 GGATTTTACAGT 213

Query Match          100.0%; Score 12; DB 14; Length 228;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatttcacagt 12
|||||
Db 202 GGATTTTACAGT 213

RESULT 9
AIU38677
LOCUS      231 bp      RNA      VRL      14-FEB-1997
DEFINITION Avian infectious bronchitis virus Al211 spike protein (S1) gene,
            hypervariable region 1, partial cds.
ACCESSION  U38677
VERSION     U38677.1  GI:1055301
KEYWORDS
SOURCE     Avian infectious bronchitis virus.
            Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
            Coronaviridae; Coronavirus.
ORGANISM   1 (sites)
            Wang,C.H. and Tsai,C.T.
            Genetic grouping for the isolates of avian infectious bronchitis
            virus in Taiwan
JOURNAL   Arch. Virol. 141 (9), 1677-1688 (1996)
MEDLINE   97049060
REFERENCE  2 (bases 1 to 231)
AUTHORS   Wang,C.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vettrinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES   Location/Qualifiers
source     1..231
           /organism="Avian infectious bronchitis virus"
           /strain="A1211"
           /db_xref="taxon:11120"
gene       1..231
           /gene="S1"
           <1..>231
CDS        /gene="S1"
           /note="hypervariable region 1; N-terminus"
           /codon_start=2
           /product="spike protein"
           /protein_id="AAB47435.1"
           /db_xref="GI:1055302"
           /translation="GWLHGGAYAVVNVSSSTNNAGSACTVGTIRGDRVVNASSIA
MTAPVGGQMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT      56 a 40 c 55 g 80 t
ORIGIN

Query Match      100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 10
AU038678
LOCUS      231 bp RNA VRL 14-FEB-1997
DEFINITION Avian infectious bronchitis virus A1960 spike protein (S1) gene,
hypervariable region 1, partial cds.
ACCESSION U38678
VERSION   1
KEYWORDS
SOURCE    U38678.1 GI:1055303
ORGANISM
          Avian infectious bronchitis virus.
          Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
          Coronaviridae; Coronavirus.

REFERENCE   1 (sites)
AUTHORS    Wang,C.H. and Tsai,C.T.
TITLE      Genetic grouping for the isolates of avian infectious bronchitis
virus in Taiwan
JOURNAL    Arch. Virol. 141 (9), 1677-1688 (1996)
MEDLINE    97049060
REFERENCE   2 (bases 1 to 231)
AUTHORS    Wang,C.
TITLE      Direct Submission
JOURNAL    Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vettrinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES   Location/Qualifiers
source     1..231
           /organism="Avian infectious bronchitis virus"
           /strain="A1955"
           /db_xref="taxon:11120"
gene       1..231
           /gene="S1"
           <1..>231
CDS        /gene="S1"
           /note="hypervariable region 1; N-terminus"
           /codon_start=2
           /product="spike protein"
           /protein_id="AAB47439.1"
           /db_xref="GI:1055310"
           /translation="GWLHGGAYAVVNVSSSTNNAGTAGCTVGIISGDRVVNASSIA
MSAPVGGQMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT      57 a 40 c 55 g 79 t
ORIGIN

Query Match      100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 12
AU025328
LOCUS      285 bp DNA STS 02-MAR-1999
DEFINITION Rattus norvegicus, OTSUKA clone, OT57.23/887g03, microsatellite
sequence, sequence tagged site.
ACCESSION AU025328
VERSION   1
KEYWORDS
SOURCE    AU025328.1 GI:4515251
          Rattus norvegicus DNA, clone:OT57.23/887g03.
          Rattus norvegicus
ORGANISM
```

```

TITLE      Direct Submission
JOURNAL    Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vettrinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES   Location/Qualifiers
source     1..231
           /organism="Avian infectious bronchitis virus"
           /strain="A1211"
           /db_xref="taxon:11120"
gene       1..231
           /gene="S1"
           <1..>231
CDS        /gene="S1"
           /note="hypervariable region 1; N-terminus"
           /codon_start=2
           /product="spike protein"
           /protein_id="AAB47435.1"
           /db_xref="GI:1055302"
           /translation="GWLHGGAYAVVNVSSSTNNAGSACTVGTIRGDRVVNASSIA
MTAPVGGQMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT      56 a 40 c 55 g 80 t
ORIGIN

Query Match      100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 11
AU038681
LOCUS      231 bp RNA VRL 14-FEB-1997
DEFINITION Avian infectious bronchitis virus A1955 spike protein (S1) gene,
hypervariable region 1, partial cds.
ACCESSION U38681
VERSION   1
KEYWORDS
SOURCE    U38681.1 GI:1055309
          Avian infectious bronchitis virus.
          Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
          Coronaviridae; Coronavirus.

REFERENCE   1 (sites)
AUTHORS    Wang,C.H. and Tsai,C.T.
TITLE      Genetic grouping for the isolates of avian infectious bronchitis
virus in Taiwan
JOURNAL    Arch. Virol. 141 (9), 1677-1688 (1996)
MEDLINE    97049060
REFERENCE   2 (bases 1 to 231)
AUTHORS    Wang,C.
TITLE      Direct Submission
JOURNAL    Submitted (17-OCT-1995) Ching-Ho Wang, National Taiwan University,
Vettrinary Medicine, 142 Chou-san Rd., Taipei, 106, Taiwan
FEATURES   Location/Qualifiers
source     1..231
           /organism="Avian infectious bronchitis virus"
           /strain="A1955"
           /db_xref="taxon:11120"
gene       1..231
           /gene="S1"
           <1..>231
CDS        /gene="S1"
           /note="hypervariable region 1; N-terminus"
           /codon_start=2
           /product="spike protein"
           /protein_id="AAB47439.1"
           /db_xref="GI:1055310"
           /translation="GWLHGGAYAVVNVSSSTNNAGTAGCTVGIISGDRVVNASSIA
MSAPVGGQMQWSKQFCTAHCNFSDFTVFVTH"
BASE COUNT      57 a 40 c 55 g 79 t
ORIGIN

Query Match      100.0%; Score 12; DB 14; Length 231;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 205 GGATTTTACAGT 216

RESULT 12
AU025328
LOCUS      285 bp DNA STS 02-MAR-1999
DEFINITION Rattus norvegicus, OTSUKA clone, OT57.23/887g03, microsatellite
sequence, sequence tagged site.
ACCESSION AU025328
VERSION   1
KEYWORDS
SOURCE    AU025328.1 GI:4515251
          Rattus norvegicus DNA, clone:OT57.23/887g03.
          Rattus norvegicus
ORGANISM
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE AUTHORS

Watanabe, T.K., Hishigaki, H., Kanemoto, N., Miyakita-Mizoguchi, A., Oga, K., Okuno, S., Ono, T., Tsuji, A., Hayashi, H., Adachi, M., Yamasaki, Y., Iriye, Y., Takahashi, E., Takagi, T., Nakamura, Y. and Tanigami, A.

The large-scale mapping of rat microsatellite markers

JOURNAL REFERENCE

Unpublished (1998)

REFERENCE AUTHORS

Watanabe, T.K.

JOURNAL TITLE

Submitted (24-JUL-1998) to the DDBJ/EMBL/GenBank databases. Takeshi K Watanabe, Otsuka Pharmaceutical Co., Ltd., Otsuka GEN Research Institute, 463-10, Kagasuno, Kawachi-cho, Tokushima, Tokushima 771-0192, Japan (E-mail: watanabe@otsuka.genome.ad.jp, Tel: 81-886-65-2888, Fax: 81-886-37-1035)

FEATURES source

1. .285

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="OF57.23/887g03"

/notes="OF57.23/887g03P-5'-CCTTATTCCATGGGTACCTGA-3',

OF57.23/887g03R-5'-CGATCAGGTACTGCCATACC-3"

114 a 80 c 35 g 56 t

BASE COUNT ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 285;
Best Local Similarity 100.0%; Pred. No. 5e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 205 GGATTTTACAGT 216

RESULT 13 HPCTH13P30

LOCUS

Hepatitis C virus RNA for NS5, partial sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

gene

CDS

/gene="TH13PI"
/codon_start=1
/product="NS5"
/protein_id="BAA05996.1"
/db_xref="GI:1669629"
/translation="TPVNSWLGNIIMYAPTINVRVLMTHFFGILQSQETLHKALDFD
MTGVYSITPLDLPLQIIQRKHGMAAFSLHGISPGELNVRVAASLRKLGAPPLRAWRHRA
RAVRRAKLQAQGKAATCGRY"

BASE COUNT 89 a 100 c 94 g 83 t

ORIGIN

Query Match 100.0%; Score 12; DB 14; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 86 GGATTTTACAGT 97

RESULT 14

IBR243261

LOCUS

IBR243261

368 bp RNA

VRL

31-JAN-2000

DEFINITION

Infectious bronchitis virus S gene, strain 624/I (368bp).

ACCESSION

AJ243261

VERSION

AJ243261.1

GI:5262733

KEYWORDS

S gene.

SOURCE

Avian infectious bronchitis virus.

ORGANISM

Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

Coronaviridae; Coronavirus.

REFERENCE

1 (bases 1 to 368)

AUTHORS

Capua, I., Minta, Z., Karpinska, E., Mawditt, K., Britton, P.,

Cavanagh, D. and Gough, R.E.

TITLE

Co-circulation of four types of infectious bronchitis virus (793/B,

624/I, B1648 and Massachusetts)

JOURNAL

Avian Pathol. 28, 587-592 (1999)

REFERENCE

2 (bases 1 to 368)

AUTHORS

Capua, I.

TITLE

Direct Submission

JOURNAL

Submitted (18-JUN-1999) Capua I., Istituto Zooprofilattico delle

Venezie, Via Romea 14/A, 35020, Legnaro, Padova, Italy

FEATURES

Location/Qualifiers

1. .368

/organism="Avian infectious bronchitis virus"

/strain="624/I"

/db_xref="taxon:11120"

gene

1. .368

/gene="S"

CDS

<1. .>368

/gene="S"

/codon_start=1

/protein_id="CAB45685.1"

/db_xref="GI:5262734"

/db_xref="SPTREMBL:Q9WML2"

/translation="NAGTAQCCTAGTAISMSKNSFSSAVMTAPGSGMQMSTNQFCTAH

CNFTDTFTVTHCFKSGAGQCPLTGTFLPSGVIRVSAMRKGNFSLFYNLTVSVTKYPTF

KSLQCVNHNHVSFLNGDLVFT"

BASE COUNT 92 a 73 c 75 g 128 t

ORIGIN

Query Match 100.0%; Score 12; DB 14; Length 368;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12

|||||

Db 144 GGATTTTACAGT 155

RESULT 15

G30656
LOCUS 388 bp DNA STS 05-OCT-1996
DEFINITION human STS SHGC-37261, sequence tagged site.
ACCESSION G30656
VERSION G30656.1 GI:1594207
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 388)
AUTHORS Myers,R.M.
JOURNAL Unpublished (1996)
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu
Primer A: TAAGCAACACCACTAATGTCCTGG
Primer B: CTTGCAGTTATGCTAGATGGG
STS size: 128
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds
Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 10 ul
Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3
Prepared with primer pairs provided by Sandoz, derived from H66941
-- Washington University/Merck EST sequence.
FEATURES
source 1..388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13"
STS 10..137
primer_bind 10..33
primer_bind complement(115..137)
BASE COUNT 130 a 74 c 68 g 116 t
ORIGIN
Query Match 100.0%; Score 12; DB 11; Length 388;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggattttacagt 12
|||||
Db 152 GGATTTTACAGT 163

Search completed: April 3, 2002, 05:13:00
Job time: 4095 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:39:12 ; Search time 2120.21 seconds
(without alignments)
471.348 Million cell updates/sec

Title: US-09-700-187-2

Perfect score: 93

Sequence: 1 aaaagtaacacatatattga.....ttacagtaataagaacga 93

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpi:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
c 1	37.8	40.6	357	AA169463	AA169463 z084h08.r
c 2	36.4	39.1	320	BH036705	BH036705 RPCI-24-2
c 3	36.4	39.1	555	BH102413	BH102413 RPCI-24-2
c 4	35.4	38.1	489	CNS03DHC	AL239097 Tetraodon
c 5	35.4	38.1	1714	BG024018	BG024018 602303237
c 6	34.8	37.4	521	C90222	C90222 C90222 Dict
c 7	34.6	37.2	631	AV405634	AV405634 AV405634
c 8	34.4	37.0	442	AQ776172	AQ776172 HS_5571_A
c 9	34	36.6	846	AQ743241	AQ743241 HS_5388_B
c 10	33.6	36.1	515	AQ764101	AQ764101 HS_2001_A
c 11	33.4	35.9	587	AJ388941	AJ388941 AJ388941
c 12	33.2	35.7	545	AQ545066	AQ545066 CITBI-EL-

c 13	33	35.5	545	10	AV739051	AV739051
c 14	33	35.5	547	13	AZ648187	AZ648187 1M0517001
c 15	33	35.5	549	13	TA195A02P	TA195A02P
c 16	33	35.5	558	13	TA12F02P	TA12F02P
c 17	33	35.5	947	11	BG025969	BG025969 602292131
c 18	33	35.5	1082	13	CNS05FLS	AL335161 Tetraodon
c 19	32.8	35.3	321	13	B56628	B56628 CIT-HSP-200
c 20	32.6	35.1	924	13	B12729	B12729 F3H9-Sp6 IG
c 21	32.6	35.1	1190	13	B10407	B10407 F20B22-T7 I
c 22	32.4	34.8	223	10	BE030084	BE030084 kp38f02.y
c 23	32.4	34.8	631	13	AZ334849	AZ334849 1M0064124
c 24	32.4	34.8	669	13	AZ516773	AZ516773 RPCI-11-4
c 25	32.4	34.8	1062	13	CNS00LU6	AL078754 Drosophila
c 26	32.2	34.6	964	13	CNS005OV	AL060482 Drosophila
c 27	32	34.4	534	11	BG555297	BG555297 df01h08.x
c 28	32	34.4	907	11	BE873019	BE873019 601451032
c 29	31.8	34.2	323	10	AV017830	AV017830 AV017830
c 30	31.8	34.2	596	10	BE217602	BE217602 MD0232 Me
c 31	31.8	34.2	945	13	CNS0605W	AL407658 T7 end of
c 32	31.8	34.2	1099	10	AL536986	AL536986 AL536986
c 33	31.8	34.2	1101	13	CNS016VF	AL107253 Drosophila
c 34	31.6	34.0	216	10	AL699725	AL699725 tzl1c02.x
c 35	31.6	34.0	237	11	F02028	F02028 HSCMD022 n
c 36	31.6	34.0	252	11	N73114	N73114 yv69b11.r1
c 37	31.6	34.0	343	10	AI038703	AI038703 ox35d05.s
c 38	31.6	34.0	346	11	R92539	R92539 yq08d05.r1
c 39	31.6	34.0	350	10	AW512010	AW512010 xu75h01.x
c 40	31.6	34.0	356	10	AI200734	AI200734 qf55f12.x
c 41	31.6	34.0	370	10	AW504912	AW504912 UI-HF-BN0
c 42	31.6	34.0	393	10	AW010162	AW010162 ST02G07 P
c 43	31.6	34.0	414	11	H47009	H47009 y016g02.s1
c 44	31.6	34.0	417	11	R52980	R52980 y985a10.s1
c 45	31.6	34.0	420	11	N55097	N55097 yv43f04.s1

ALIGNMENTS

RESULT 1
AA169463/c

LOCUS

DEFINITION

z084h08.r1 Stratagene ovarian cancer (#937219) Homo sapiens CDNA

clone IMAGE:593631 5', mRNA sequence.

ACCESSION

AA169463.1 GI:1748411

VERSION

human.

KEYWORDS

EST.

SOURCE

Homo sapiens

REFERENCE

1 (bases 1 to 357)

Authors

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1082 Std Error: 0.00

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 251.

Location/Qualifiers

1..357

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:593631"

```

/clone_lib="Stratagene ovarian cancer (#937219)"
/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Papillary
serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector:
-5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"
BASE COUNT      104 a      48 c      57 g      146 t
ORIGIN
Query Match      40.6%; Score 37.8; DB 10; Length 357;
Best Local Similarity 64.0%; Pred. No. 37;
Matches 57; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgaataatttatactaaactatttcttagtactgttaac 60
||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 AAACAACAGCTAAGCAAAATTAATTAATAAATAAACAATTTCTAAGTAGATGATATA 157
QY 61 atgtctgaggtttacagtaataaaga 89
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 ATCACGATGTTTATACAGCTATAAGCA 128

RESULT 2
BH036705          320 bp      DNA          GSS          17-JUL-2001
LOCUS             RPCI-24-22406.TJ RPCI-24 Mus musculus genomic clone RPCI-24-22406,
DEFINITION        DNA sequence.
ACCESSION          BH036705
VERSION            BH036705.1 GI:14811846
KEYWORDS           GSS.
SOURCE             house mouse.
ORGANISM           Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 320)
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Akinret,B., Levins,M.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 224 row: 0 column: 6
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..320
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-22406"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J"

FEATURES
source

```

```

BASE COUNT      102 a      58 c      47 g      113 t
ORIGIN
Query Match      39.1%; Score 36.4; DB 13; Length 320;
Best Local Similarity 64.0%; Pred. No. 77;
Matches 55; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 aaagtaacacatatatttgaataatttatactaaactatttcttagtactgttaac 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 AAAGAATAACAATAATATGACTCCCTTTTACTTAAACAATAATTTAGTAATCTTCTC 218
QY 61 atgtctgaggtttacagtaataaa 86
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 TTCCTTAAGTTATTAGCAGCAATAGA 244

RESULT 3
BH102413          555 bp      DNA          GSS          19-JUL-2001
LOCUS             RPCI-24-229N23.TJ RPCI-24 Mus musculus genomic clone RPCI-24-229N23
DEFINITION        DNA sequence.
ACCESSION          BH102413
VERSION            BH102413.1 GI:14928295
KEYWORDS           GSS.
SOURCE             house mouse.
ORGANISM           Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 555)
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 229 row: N column: 23
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..555
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-229N23"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT      143 a      113 c      86 g      213 t
ORIGIN
Query Match      39.1%; Score 36.4; DB 13; Length 555;
Best Local Similarity 64.0%; Pred. No. 66;
Matches 55; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```



```

Qy 1 aaaaagtaacacatatatttgataaaatttattactaaaactattttctagtagtactgttaatc 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 AAAGAATAACAATAATGACTCCCTTTTACTTAACAATAATTTTACTAATCTCTTC 215

Qy 61 atgtctgaggattttacagtaataaa 86
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 216 TTCCTTAAGTTATTAGCAGCAATAGA 241

RESULT 4
CNS03DHC/c
LOCUS 489 bp DNA GSS 15-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ORI end of clone
017A03 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL239097.1 GI:7898232
VERSION AL239097.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 489)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fzimes,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 489)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fzimes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 489)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
Location/Qualifiers
1..489
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="017A03"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG017AA02SP1-end :
PUC-ORI"
BASE COUNT 243 a 21 c 105 g 85 t 35 others
ORIGIN

Query Match 38.1%; Score 35.4; DB 13; Length 489;
Best Local Similarity 64.6%; Pred. No. 1.1e+02;
Matches 51; Conservative 1; Mismatches 27; Indels 0; Gaps 0;

Qy 6 taacacatatttataaaatttactaaaactattttctagtagtactgttaatc 65
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 TTAACAATTTTAAACAATCTCACWAAATCTTTTCTCTTCTCTTATTTTC 38

Qy 66 tgaggattttacagtaata 84
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 ATCAAAATTTAAATAATA 19

RESULT 5

```

```

BG024018
LOCUS 1714 bp mRNA EST 24-JAN-2001
DEFINITION BG02303237F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4394600 5',
mRNA sequence.
ACCESSION BG024018
VERSION BG024018.1 GI:12409159
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1714)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10090 row: j column: 09
High quality sequence stop: 30.
FEATURES
Location/Qualifiers
1..1714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4394600"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 Kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 919 a 209 c 457 g 129 t
ORIGIN

```

```

Query Match 38.1%; Score 35.4; DB 11; Length 1714;
Best Local Similarity 61.3%; Pred. No. 80;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1 aaaagtaacacatatatttgataaaatttactaaaactattttctagtagtactgttaatc 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 AAATAATAATATATATTATTATAATATATATAATATATATAATATATATATATATTTT 125

Qy 61 atgtctgaggattttacagtaataaaagaaacga 93
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 126 TTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 158

RESULT 6
LOCUS 521 bp mRNA EST 20-APR-1998
DEFINITION C90222 Dictyostelium discoideum SS (H.Urushi-hara) Dictyostelium
discoideum cDNA clone SSI395, mRNA sequence.
ACCESSION C90222
VERSION C90222.1 GI:3059842
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
1 (bases 1 to 521)
Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
Developmental cDNA in Dictyostelium discoideum
Unpublished (1998)
TITLE
JOURNAL

```

```

COMMENT      Contact: Hideko Urushihara
              Institute of Biological Sciences
              University of Tsukuba
              3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
              Email: d402huesakura.cc.tsukuba.ac.jp.

FEATURES
source
1..521
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSI395"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT   253 a 59 c 73 g 135 t 1 others
ORIGIN

Query Match   37.4%; Score 34.8; DB 11; Length 521;
Best Local Similarity 65.4%; Pred. No. 1.5e+02;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 13 tatttgataaattattactaaactattttctagctactgttaacatgctgagat 72
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 TATAATAAATTTTATGCTATAATTTAATAATAATAATAATAATAATAATAAT 478
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 73 ttacagtaataaagaa 90
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 TTAATAATAATAAAAAA 496

RESULT 7
AV405634
LOCUS
DEFINITION   AV405634 Bombyx mori wing disk C108 5th-instar day-3 larva Bombyx
              mori cDNA clone wdv30047 T3, mRNA sequence.
ACCESSION   AV405634
VERSION     AV405634.1
KEYWORDS    EST.
SOURCE      AV405634.1 GI:6909722
            domestic silkworm.
ORGANISM    Bombyx mori
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
            ; Bombycoidea; Bombycidae; Bombyx.
REFERENCE   1 (bases 1 to 631)
AUTHORS     Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE       Bombyx mori cDNA
JOURNAL     Unpublished (2000)
COMMENT     Contact: Mita K
            Genome Research Group
            National Institute of Radiological Sciences
            Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
            Email: kmita@nirs.go.jp
            method: uni-directional, sequence direction: sequenced from T3 primer
            (5' -> 3')

FEATURES
source
1..631
/organism="Bombyx mori"
/strain="C108"
/db_xref="taxon:7091"
/clone="wdv30047"
/clone_lib="Bombyx mori wing disk C108 5th-instar day-3
            larva"
/sex="female/male mixed"
/tissue_type="wing disk"
/dev_stage="5th-instar day-3 larva"

BASE COUNT   238 a 94 c 68 g 231 t
ORIGIN

Query Match   37.2%; Score 34.6; DB 10; Length 631;
Best Local Similarity 63.1%; Pred. No. 1.9e+02;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 8 acacataatttgataaattattactaaactattttctagctactgttaacatgctg 67
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 ACACATATATACACATATATATTTTACATATATATATATATATATATATATATGCTG 349
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Best Local Similarity 61.8%; Pred. No. 1.6e+02;
Matches 55; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 aaagtaaacacacataatttgataaattattactaaactattttctagctactgttaac 60
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 AATAGAACATTTATTTTGGTATAGTAATACACAGACAATTTTATATATATATTTATT 148
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 atgtctgaggattttacagtaataaagaa 89
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 AAGAATGAACATAAATAAATAAATAAATAA 177

RESULT 8
AQ776172/c
LOCUS
DEFINITION   AQ776172 442 bp DNA GSS
              HS_5571_A2_G04_T7A RPCI-11 Human Male BAC Library Homo sapiens
              genomic clone Plate-1147 Col=8 Row=M, DNA sequence.
ACCESSION   AQ776172
VERSION     AQ776172.1
KEYWORDS    GSS.
SOURCE      AQ776172.1 GI:5655900
            human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 442)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            plate: 1147 row: M column: 8
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 442.
FEATURES
source
1..442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate-1147 Col=8 Row=M"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
            Male blood DNA was isolated from one randomly chosen donor
            and partially digested with a combination of EcoRI and
            EcoRI Methylase. Size selected DNA was cloned into the
            pBACE3.6 vector at EcoRI sites"

BASE COUNT   119 a 79 c 69 g 170 t
ORIGIN

Query Match   37.0%; Score 34.4; DB 13; Length 442;
Best Local Similarity 63.1%; Pred. No. 1.9e+02;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 8 acacataatttgataaattattactaaactattttctagctactgttaacatgctg 67
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 ACACATATATACACATATATATTTTACATATATATATATATATATATATATGCTG 349
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy 68 aggattttacagtaataaagaac 91
||| ||| ||| ||| ||| |||
Db 348 AGTAATATGCAGCCATAAACACAC 325

RESULT	9
LOCUS	AQ743241/c
DEFINITION	AQ743241 846 bp DNA 16-JUL-1999 HS_5388_B1_D10.SP6 RPI-CI-11 Human Male BAC Library Homo sapiens genomic clone Plate-964 Col-19 Row-H, DNA sequence.
ACCESSION	AQ743241
VERSION	AQ743241.1 GI:5520763
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 846)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequencing-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887

RESULT 10
AQ764101

LOCUS	AQ764101	515 bp	DNA	GSS	28-JUL-1999
DEFINITION	HS_2001_AI_G12_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2001 Col=23 Row=M, DNA sequence.				
ACCESSION	AQ764101				
VERSION	AQ764101.1	GI:5642217			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 515) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)				
MEDLINE	99380589				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end web server: http://www.htsc.washington.edu Plate: 2001 row: M column: 23 Seq primer: M13 Reverse Class: BAC ends High quality sequence stop: 515.				
FEATURES	source				
	Location/Qualifiers				
	1..515				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="Plate=2001 Col=23 Row=M"				
	/clone_lib="CIT Approved Human Genomic Sperm Library D"				
	/sex="male"				
	/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"				
BASE COUNT	178 a	84 c	86 g	156 t	11 others
ORIGIN					
	Query Match	36.1%	Score 33.6;	DB 13;	Length 515;
	Best Local Similarity	61.4%;	Pred. No. 2.7e+02;		
	Matches	54;	Conservative	0;	Mismatches 34; Indels 0; Gaps 0;
QY	2 aaagtaacacacatatttgataaatttataactaaacattttctagctactgttaatca 61				
Db	49 ACATTCAACATACATACATAGATTTGTGTCAAACATTTGACTAGTACTAATAAATA 108				
QY	62 tgtctgaggattttacagtaataaagaa 89				
Db	109 ATTGTGTGATTGATATAGTAGACAGTA 136				
RESULT	11				
AJ388941/C					
LOCUS	AJ388941	587 bp	mRNA	EST	21-MAR-2000
DEFINITION	AJ388941 Medicago truncatula R108 Medicago truncatula cDNA clone M1N0312, mRNA sequence.				
ACCESSION	AJ388941				
VERSION	AJ388941				
KEYWORDS	AJ388941.1 GI:6604208				
SOURCE	EST.				
ORGANISM	barrel medic.				
	Medicago truncatula				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.				
REFERENCE	1 (bases 1 to 587)				
AUTHORS	Goryunov,J.,Vaubert,D., Jimenez-Zurdo,J.I., Charon,C., Troussard				

TITLE L., Kondorosi A. and Kondorosi, F.
JOURNAL Analysis of *Medicago truncatula* module expressed sequence tags
MEDLINE Mol. Plant Microbe Interact. 13 (1), 62-71 (2000)
COMMENT 20120379
Contact: Gyorgyey J
Institut des Sciences Vegetales
Centre National de la Recherche Scientifique (CNRS)
Av. de la Terrasse Bat.23, Gif-sur-Yvette, 91198, FRANCE
Additional information about these EST clones can be downloaded
from
<http://www.cnrs-gif.fr/isv/AK/index.html>. (URL provided by the
author).

FEATURES	SOURCE
----------	--------

FEATURES	source	Location/Qualifiers
1.	.587	/organism="Medicago truncatula"
		/strain="R108"
		/db_xref="taxon:3880"
		/clone="MtN0312"
		/clone_lib="Medicago truncatula R108"
		/tissue_type="symbiotic root nodule"
		/dev_stage="developing young nodule"
		/notes="Vector: NOL1y Lambda HybriZAP"
BASE COUNT	201 a	75 c 99 g 206 t 6 others

BASE COUNT	201 a	75 c	99 g	206 t	6 others
ORIGIN					
Query Match		35.9%	Score 33.4;	DB 10;	Length 587;
Best Local Similarity		64.5%;	Pred. No. 2.9e+02;		
Matches 49; Conservative		0;	Mismatches 27;	Indels 0;	Gaps 0;

Qy. 3 aagtaacacacatttttgataaaattttttactactaaacacattttctagtagtactttgttaatacat 62
 447 AAATTTCAAAAATTTTCATCTTATTATTAATAATATTTCTAGTCTCGTAAATCAT 388
 Qy 63 gtctgaggattttaca 78
 Db 387 ACATAAGGACTTTATA 372

RESULT	12
AQ545066/C	
LOCUS	
DEFINITION	AQ545066 545 bp DNA GSS 28-MAY-1999 CITR1-EI - 263409.TF CITR1-E1 Homo sapiens genomic clone 263409, DNA sequence.
ACCESSION	AQ545066
VERSION	AQ545066.1 GI:4903792
KEYWORDS	GSS.
SOURCE	human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 545)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbswtigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21

```

FEATURES
source      Location/Qualifiers
1..545     /organism="Homo sapiens"
           /db xref="taxon:9606"

```

```

/clone="263409"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
185 a 96 c 87 g 177 t

```

```
Query Match          35.7%; Score 33.2; DB 13; Length 545;
Best Local Similarity 61.6%; Pred. NO. 3.3e+02;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
```

Qy 2 aaagtaacacatatatttgataaattattactaaactatttctctagtacttctgtaataca 61
Db 184 ATAGTGAATTATTTTCTCGAAATTTCTGAAAAATAGTCAATAATTAATGACAATCG 125
Qy 62 tgtctgaggattttacagtaataaag 87
Db 124 AGTCTGAATAAAATAGCATATAAAG 99

RESULT	13
LOCUS	AV739051/C
DEFINITION	545 bp mRNA EST
ACCESSION	AV739051 CB Homo sapiens CDNA clone CBLAPG08 5', mRNA sequence.
VERSION	AV739051.1 GI:10856632
KEYWORDS	EST.
SOURCE	human.

ORGANISM	Homo sapiens
Source	Human
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Shanghai Institute of Hematology, Rui-Jin Hospital
197 Rui-Jin II Road, Shanghai 200025, P. R. China
Tel: 86-21-64740490
Fax: 86-21-64743206
Email: mbshi@ems.stn.sh.cn
This clone is available at Shanghai Hematology Institute in
Shanghai.

Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

FEATURES
source
1. .545
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CBLAPG08"
/clone_lib="CB"
/tissue_type="cord blood"
/cell_type="CD34+ hemato
/lab_host="BM25.8"
/note="Vector: pBluescri
cloned randomly with the
171 a 118 c 124 g 13
BASE COUNT
```

Query Match 35.5%; Score 33; DB 10; Length 545;
Best Local Similarity 60.7%; Pred. NO. 3.6e+02;

Qy **2** aaagtaacacataatttgatataaaatttactaaaactttttctagtactgtttaaca 61
||| | | ||| | |||| | | | | | | |
Db **210** AAATTCTGATAAATTCGTAAATAAGTCTCAGGAGAACCACAAAATAGACTTTAGTCAATCA 151

Qy **62** tgctdaggaatttacagaataaaaaaa 90

```

Db 150 TGTGGAAACTTAAGAGTAATGAAGTAA 122
|||||
RESULT 14
AZ648187 547 bp DNA GSS 14-DEC-2000
LOCUS IM0517001F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION clone UUGCLM0517001 F, DNA sequence.
ACCESSION AZ648187
VERSION AZ648187.1 GI:11780403
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 547)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mamoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0517 row: 0 column: 01
Seq primer: CGTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 547.
FEATURES
source
1. 547
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 209 a 86 c 67 g 185 t
ORIGIN

Query Match 35.5%; Score 33; DB 13; Length 547;
Best Local Similarity 60.7%; Pred. No. 3.6e+02;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Db 150 TGTGGAAACTTAAGAGTAATGAAGTAA 122
|||||
RESULT 14
AZ648187 547 bp DNA GSS 14-DEC-2000
LOCUS IM0517001F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION clone UUGCLM0517001 F, DNA sequence.
ACCESSION AZ648187
VERSION AZ648187.1 GI:11780403
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 547)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mamoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0517 row: 0 column: 01
Seq primer: CGTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 547.
FEATURES
source
1. 547
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 209 a 86 c 67 g 185 t
ORIGIN

```

```

Qy 1 aaaagtaacacataatttggataaatttattactaaacacttttctagtaactgttaacc 60
|||||
Db 378 AACGCAATATATATATACATACAAATTACTATTAAAGTTTATTTCAGGCAATTTCTTATG 437
|||||
Qy 61 atgtctgaggattttacagtaataaagaa 89
|||||
Db 438 TGTCTTGAGGCTGCTTAGGAAATTTAGAA 466
|||||
RESULT 15
TA195A02P/c TA195A02P 549 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 195a02, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL477905
VERSION AL477905.1 GI:11841717
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
AUTHORS Trypanosoma.
1 (bases 1 to 549)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TRED927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
source
1. 549
/organism="Trypanosoma brucei"
/strain="TRED927"
/db_xref="taxon:5691"
/clone="195a02"
BASE COUNT 226 a 66 c 97 g 160 t
ORIGIN

Query Match 35.5%; Score 33; DB 13; Length 549;
Best Local Similarity 60.7%; Pred. No. 3.6e+02;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 2 aaagtaacacataatttggataaatttattactaaacacttttctagtaactgttaacca 61
|||||
Db 340 AAATATATATATTTAGGAAATATCTAAAAAGCAAAATTTTCATATATGATAAAAA 281
|||||
Qy 62 tgtctgaggattttacagtaataaagaa 90
|||||
Db 280 TTCTTAAATGTTACCAACTCAGAA 252
|||||

Search completed: April 3, 2002, 04:39:16
Job time: 2836 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:30:10 ; Search time 189.71 Seconds
(without alignments)
54.230 Million cell updates/sec

Title: us-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

1:	/SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:	*
2:	/SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:	*
3:	/SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:	*
4:	/SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:	*
5:	/SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:	*
6:	/SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:	*
7:	/SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:	*
8:	/SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:	*
9:	/SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:	*
10:	/SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:	*
11:	/SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:	*
12:	/SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:	*
13:	/SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:	*
14:	/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:	*
15:	/SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:	*
16:	/SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:	*
17:	/SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:	*
18:	/SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:	*
19:	/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:	*
20:	/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:	*
21:	/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:	*
22:	/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:	*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	12	21	AAA97383
c	2	12	100.0	22	20 AAX79146
3	12	100.0	25	21	AAA68434
4	12	100.0	30	21	AAA97393
5	12	100.0	31	21	AAA97411
c	6	12	100.0	31	21 AAA97412
7	12	100.0	31	21	AAA97423
8	12	100.0	31	21	AAA97424
9	12	100.0	31	21	AAA97427
10	12	100.0	31	21	AAA97428
11	12	100.0	31	22	AA130154

C	12	12	100.0	39	21	AA97403	Pea pra2 light-rep
	13	12	100.0	48	21	AA97396	Pea pra2 light-rep
	14	12	100.0	52	21	AA97397	Pea pra2 light-rep
C	15	12	100.0	53	21	AA97418	Pea wild-type pra2
	16	12	100.0	54	21	AA97417	Pea wild-type pra2
	17	12	100.0	93	21	AA97384	Pea pra2 gene high
C	18	12	100.0	147	16	AA722540	Human gene signatu
C	19	12	100.0	300	21	AA01517	Human colon cancer
C	20	12	100.0	344	20	AAV0257	EST clone DG279.
C	21	12	100.0	468	22	AAI12374	Probe #2307 for ge
C	22	12	100.0	468	22	AAI02286	Probe #2277 used t
	23	12	100.0	547	21	AA93808	Cat flea hindgut a
C	24	12	100.0	646	21	AA256749	Human transmembran
	25	12	100.0	651	22	AAH06265	Human cDNA clone (
	26	12	100.0	724	22	AAH71534	Human cervical can
	27	12	100.0	727	21	AA44649	Arabidopsis thalia
	28	12	100.0	852	22	AAH73365	Human cervical can
C	29	12	100.0	1126	21	AA42447	Arabidopsis thalia
C	30	12	100.0	1129	20	AA27340	Human secreted pro
C	31	12	100.0	1138	19	AAV46314	Human secreted pro
C	32	12	100.0	1138	22	AA98480	Human cDNA clone B
C	33	12	100.0	1188	20	AA52916	Human prostate tum
	34	12	100.0	1221	18	AAV74319	Staphylococcus aur
	35	12	100.0	1279	21	AA34384	Arabidopsis thalia
C	36	12	100.0	1287	20	AA37458	Human secreted pro
	37	12	100.0	1341	22	AA87638	Arabidopsis thalia
C	38	12	100.0	1342	22	AAH29817	S cerevisiae apopt
C	39	12	100.0	1434	20	AA16007	DNA encoding an an
C	40	12	100.0	1466	21	AA40161	Arabidopsis thalia
C	41	12	100.0	1505	21	AA33544	Arabidopsis thalia
C	42	12	100.0	1605	21	AA39336	Arabidopsis thalia
	43	12	100.0	1666	21	AA77904	Human cancer assoc
	44	12	100.0	1677	22	AAH34524	Human colon cancer
	45	12	100.0	1706	22	AA98716	Human late stage o

ALIGNMENTS

RESULT 1
AAA97383
ID AAA97383 standard; DNA; 12 BP.
XX AAA97383;
XX
XX
XX 29-JAN-2001 (first entry)
XX
XX Pea pra2 gene light-repressible promoter core region.
XX
XX GTP-binding protein pra2; pea; light-repressible promoter;
XX photoinhibitory; expression cassette; transgenic plant;
XX deterioration prevention; storage; core region; ds.
XX
XX Pisum sativum.
XX
XX WO200055313-AL.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
XX under photoinhibitory conditions, and for transforming a plant cell or
XX plant to improve quality and prevent deterioration during storage -

```
PS Claim 1; Page 17; 49pp; Japanese.
XX
CC The invention relates to a light-repressible promoter (AAAG7385), or
CC active fragments thereof (AAAG7383, AAAG7384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. The present sequence
CC represents the pea pra2 promoter core region.
XX
SQ Sequence 12 BP; 3 A; 1 C; 3 G; 5 T; 0 other;

Query Match      100.0%; Score 12; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 1 ggattttacagt 12
   |||||
RESULT 2
AAAX79146/C
ID AAX79146 standard; DNA; 22 BP.
XX
AC AAX79146;
XX
DT 17-AUG-1999 (first entry)
XX
DE Primer NGA249-R for A.thaliana SSCP marker.
XX
KW MSH6; MutS homologue; plant; DNA mismatch repair; genetic variation;
KW characteristic; microsatellite; primer; PCR; amplification; SSCP; ss;
KW simple sequence length polymorphism.
XX
OS Synthetic.
OS Arabidopsis thaliana.
XX
PN WO9919492-A2.
XX
PD 22-APR-1999.
XX
PF 09-OCT-1998; 98WO-EP06977.
XX
PR 10-OCT-1997; 97AU-0009745.
XX
PA (RHON ) RHONE-POULENC AGROCHIMIE.
XX
PI Betner AS, Doutriaux M, Freyssinet G, Perez P;
XX WPI; 1999-277644/23.
XX
PT DNA encoding protein functionally involved in the DNA mismatch
PT repair system of a plant
XX
PS Example 3; Page 28; 117pp; English.
XX
CC The invention relates to the isolation of the Arabidopsis thaliana MSH3
CC (AAX79066) and MSH6 (AAX79067) genes. These genes are MutS homologues
CC (MSH) from plants and are involved in DNA mismatch repair. The DNA
CC sequence can be used in processes for at least partially inactivating a
CC DNA mismatch repair system of a plant, for increasing genetic variation
CC in a plant, and for obtaining a plant with a desired characteristic.
CC Primers AAX79105-X79160 represent 28 primer pairs used to amplify short
CC allelic repeat fragments designated Simple Sequence Length Polymorphisms
CC (SSLP). These fragments can be used as markers in the analysis of
CC homologous recombination between genomes of A.thaliana subspecies.
XX

SQ Sequence 22 BP; 7 A; 7 C; 3 G; 5 T; 0 other;

Query Match      100.0%; Score 12; DB 20; Length 22;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 21 GGATTTTACAGT 10
   |||||
RESULT 3
AAA68434
ID AAA68434 standard; DNA; 25 BP.
XX
AC AAA68434;
XX
DT 27-OCT-2000 (first entry)
XX
DE Bacteriophage 3A ORF RBS sequence 3AORF180.
XX
KW Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW bacterial growth inhibition; PCR primer; RBS; ribosome binding site;
KW bacterial infection; ss.
XX
OS Bacteriophage 3A.
XX
PN WO200032825-A2.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-IB02040.
XX
PR 03-DEC-1998; 98US-0110992.
XX
PR 03-JUN-1999; 99US-0326144.
XX
PR 28-SEP-1999; 99US-0407804.
XX
PR 30-SEP-1999; 99US-0157218.
XX
PR 01-DEC-1999; 99US-0168777.
XX
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
XX WPI; 2000-412361/35.
XX
PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX
PS Disclosure; Page 186; 456pp; English.
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.
XX
SQ Sequence 25 BP; 5 A; 4 C; 5 G; 11 T; 0 other;

Query Match      100.0%; Score 12; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
   |||||
```



```

Db 10 ggattttacagt 21
RESULT 4
AAA97393
ID AAA97393 standard; DNA; 30 BP.
XX
AC AAA97393;
XX
DT 29-JAN-2001 (first entry)
XX
DE Pea pra2 light-repressible promoter mutant Ls1 PCR primer, SEQ ID NO:11.
XX
XX GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; mutagenic PCR primer; ss.
XX
OS Pisum sativum.
OS Synthetic.
XX
XX WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -
XX
XX Example 3; Page 13; 49pp; Japanese.
XX
XX The invention relates to a light-repressible promoter (AAA97385), or
CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. Sequences AAA97393-A97398
CC represent PCR primers used in an exemplification of the invention
CC to mutate the pea pra2 promoter.
XX
XX Sequence 30 BP; 11 A; 4 C; 6 G; 9 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
Db 13 ggattttacagt 24

RESULT 5
AAA97411
ID AAA97411 standard; DNA; 31 BP.
XX
XX AAA97411;
XX
DT 29-JAN-2001 (first entry)
XX
XX Pea wild-type pra2 gene light-repressible promoter oligonucleotide, wt2.
DE GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; ss.
XX
OS Pisum sativum.
XX
XX WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -
XX
XX Example 8; Page 18; 49pp; Japanese.
XX
XX The invention relates to a light-repressible promoter (AAA97385), or
CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. Sequences AAA97411-A97412
CC represent oligonucleotides used in an exemplification of the invention
CC to generate a wild-type pea pra2 promoter fragment.
XX
XX Sequence 31 BP; 13 A; 3 C; 7 G; 8 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
Db 7 ggattttacagt 18

RESULT 6
AAA97412/C
ID AAA97412 standard; DNA; 31 BP.
XX
XX AAA97412;
XX
DT 29-JAN-2001 (first entry)
XX
XX Pea wild-type pra2 gene light-repressible promoter oligonucleotide, wt2.
DE GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; ss.
XX
OS Pisum sativum.
XX
XX WO200055313-A1.
XX
XX 21-SEP-2000.
XX

```

```

PF 03-MAR-2000; 2000WO-JP01269.
XX
PR 12-MAR-1999; 99JP-0066551.
XX
PA (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
XX under photoinhibitory conditions, and for transforming a plant cell or
XX plant to improve quality and prevent deterioration during storage -
XX Example 8; Page 18; 49pp; Japanese.
XX
XX The invention relates to a light-repressible promoter (AAA97385), or
XX active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX the pea GTP-binding protein pra2. The invention also relates to an
XX expression cassette containing the pra2 promoter or its active
XX fragments for the expression of a gene under photoinhibitory or dark
XX conditions in a plant, and to transgenic plants, their descendants
XX and plant tissues comprising the expression cassette. The expression
XX cassette of the invention can be used to generate transgenic plants in
XX which deterioration during storage in the dark is prevented. This is
XX particularly useful for agricultural products. Sequences AAA97411-A97412
XX represent oligonucleotides used in an exemplification of the invention
XX to generate a wild-type pea pra2 promoter fragment.
XX
XX Sequence 31 BP; 8 A; 7 C; 3 G; 13 T; 0 other;
SQ

Query Match 100.0%; Score 12; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db 25 GGATTTTACAGT 14

RESULT 7
AAA97423
ID AAA97423 standard; DNA; 31 BP.
XX
AC AAA97423;
XX
XX 29-JAN-2001 (first entry)
XX
XX Pea wild-type pra2 gene promoter region, bases -672 to -642.
XX
XX GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; core region; ds.
XX
XX Pisum sativum.
OS Synthetic.
XX
XX WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
XX under photoinhibitory conditions, and for transforming a plant cell or
XX plant to improve quality and prevent deterioration during storage -
XX Example 7; Fig 6a; 49pp; Japanese.
XX
XX The invention relates to a light-repressible promoter (AAA97385), or
XX active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX the pea GTP-binding protein pra2. The invention also relates to an
XX expression cassette containing the pra2 promoter or its active
XX fragments for the expression of a gene under photoinhibitory or dark
XX conditions in a plant, and to transgenic plants, their descendants
XX and plant tissues comprising the expression cassette. The expression
XX cassette of the invention can be used to generate transgenic plants in
XX which deterioration during storage in the dark is prevented. This is
XX particularly useful for agricultural products. The present sequence
XX represents a wild-type pea pra2 promoter fragment (bases -672 to -642 of
XX the pra2 gene) which comprises the pra2 promoter core region (AAA97383).
XX
XX Sequence 31 BP; 13 A; 3 C; 7 G; 8 T; 0 other;
SQ

Query Match 100.0%; Score 12; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db 7 ggattttacagt 18

RESULT 8
AAA97424
ID AAA97424 standard; DNA; 31 BP.
XX
AC AAA97424;
XX
XX 29-JAN-2001 (first entry)
XX
XX Pea pra2 gene promoter region mutant Ls1, bases -672 to -642.
XX
XX GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; core region; mutant; ds.
XX
XX Pisum sativum.
OS Synthetic.
XX
XX WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
XX under photoinhibitory conditions, and for transforming a plant cell or
XX plant to improve quality and prevent deterioration during storage -
XX Example 7; Fig 6a; 49pp; Japanese.
XX
XX The invention relates to a light-repressible promoter (AAA97385), or
XX active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX the pea GTP-binding protein pra2. The invention also relates to an
XX expression cassette containing the pra2 promoter or its active
XX fragments for the expression of a gene under photoinhibitory or dark
XX conditions in a plant, and to transgenic plants, their descendants
XX and plant tissues comprising the expression cassette. The expression
XX cassette of the invention can be used to generate transgenic plants in
XX which deterioration during storage in the dark is prevented. This is
XX particularly useful for agricultural products. The present sequence
XX represents a wild-type pea pra2 promoter fragment (bases -672 to -642 of
XX the pra2 gene) which comprises the pra2 promoter core region (AAA97383).
XX
XX Sequence 31 BP; 13 A; 3 C; 7 G; 8 T; 0 other;
SQ

```

CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. Sequences AAA97424-A97429
 CC represent mutant pea pra2 promoter fragments (corresponding to bases
 CC -672 to -642 of the gene) which comprise the pra2 promoter core region.
 XX
 SQ Sequence 31 BP; 13 A; 4 C; 7 G; 7 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;
 Best Local Similarity 100.0%; Pred. No. 4.le+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 7 ggattttacagt 18

RESULT 9
 AAA97427
 ID AAA97427 standard; DNA; 31 BP.
 AC AAA97427;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Pea pra2 gene promoter region mutant LS4, bases -672 to -642.
 XX

KW GRP-binding protein pra2; pea; light-repressible promoter;
 KW photoinhibitory; expression cassette; transgenic plant;
 KW deterioration prevention; storage; core region; mutant; ds.
 XX

OS Pisum sativum.
 OS Synthetic.

PN WO200055313-A1.

XX 21-SEP-2000.

XX 03-MAR-2000; 2000WO-JP01269.

XX 12-MAR-1999; 99JP-0066551.

XX (SUNR) SUNTORY LTD.

XX Sasaki Y, Nagano Y, Inaba T;

XX WPI; 2000-587526/55.

XX New DNA fragment or promoter for expressing a target gene, specifically
 PT under photoinhibitory conditions, and for transforming a plant cell or
 PT plant to improve quality and prevent deterioration during storage -
 XX

PS Example 7; Fig 6a; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GRP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. Sequences AAA97424-A97429
 CC represent mutant pea pra2 promoter fragments (corresponding to bases
 CC -672 to -642 of the gene) which comprise the pra2 promoter core region.
 XX

SQ Sequence 31 BP; 9 A; 5 C; 9 G; 8 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;

Best Local Similarity 100.0%; Pred. No. 4.le+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ggattttacagt 12
 |||||
 Db 7 ggattttacagt 18

RESULT 10
 AAA97428
 ID AAA97428 standard; DNA; 31 BP.
 XX
 AC AAA97428;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Pea pra2 gene promoter region mutant LS5, bases -672 to -642.

KW GRP-binding protein pra2; pea; light-repressible promoter;
 KW photoinhibitory; expression cassette; transgenic plant;
 KW deterioration prevention; storage; core region; mutant; ds.
 XX

OS Pisum sativum.
 OS Synthetic.

PN WO200055313-A1.

XX 21-SEP-2000.

XX 03-MAR-2000; 2000WO-JP01269.

XX 12-MAR-1999; 99JP-0066551.

XX (SUNR) SUNTORY LTD.

XX Sasaki Y, Nagano Y, Inaba T;

XX WPI; 2000-587526/55.

XX New DNA fragment or promoter for expressing a target gene, specifically
 PT under photoinhibitory conditions, and for transforming a plant cell or
 PT plant to improve quality and prevent deterioration during storage -
 XX

PS Example 7; Fig 6a; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GRP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. Sequences AAA97424-A97429
 CC represent mutant pea pra2 promoter fragments (corresponding to bases
 CC -672 to -642 of the gene) which comprise the pra2 promoter core region.
 XX

SQ Sequence 31 BP; 11 A; 4 C; 7 G; 9 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 31;
 Best Local Similarity 100.0%; Pred. No. 4.le+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 7 ggattttacagt 18

RESULT 11
 AA130154
 ID AA130154 standard; DNA; 31 BP.

```

XX AC AAI30154;
XX XX
XX DT 18-OCT-2001 (first entry)
XX XX
XX DE Human single nucleotide polymorphism (SNP) PCCB 3.
XX XX
XX KW Human; resequence; genotype; disease; forensic; paternity testing;
XX KW single nucleotide polymorphism; SNP; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Variation replace(16,A)
XX FT /*tag= a
XX FT /standard_name= "single nucleotide polymorphism"
XX PN WO200166800-A2.
XX XX
XX PD 13-SEP-2001.
XX XX
XX PF 07-MAR-2001; 2001WO-US07268.
XX XX
XX PR 07-MAR-2000; 2000US-0187510.
XX PR 22-MAY-2000; 2000US-0206129.
XX XX
XX PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX XX
XX PI Cargill M, Ireland JS, Lander ES;
XX XX
XX DF WPI; 2001-522952/57.
XX XX
XX PT Nucleic acid molecules from the human genome which include polymorphic
XX PT sites, useful in methods for predicting the presence, absence or
XX PT severity of a particular phenotype or disorder (e.g. diabetes)
XX PT associated with a particular genotype
XX XX
XX PS Claim 1; Page 69; 145pp; English.
XX XX
XX CC The invention relates to the identification of nucleic acid molecules
XX CC (AAI29513-AAI31314) from the human genome which include polymorphic sites
XX CC which can predispose individuals to disease. Various genes from a number
XX CC of individuals were resequenced and single nucleotide polymorphisms
XX CC (SNPs) in these genes discovered. The method is useful for predicting the
XX CC presence, absence or severity of a particular phenotype or disorder (e.g.
XX CC diabetes) associated with a particular genotype. The nucleic acids
XX CC containing the polymorphic sites may be useful in forensics and paternity
XX CC testing.
XX XX
XX SQ Sequence 31 BP; 6 A; 4 C; 5 G; 16 T; 0 other;

Query Match 100.0%; Score 12; DB 22; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 16 ggattttacagt 27

RESULT 12
AAI97403/C
ID AAA97403 standard; DNA; 39 BP.
XX AC AAA97403;
XX XX
XX DT 29-JAN-2001 (first entry)
XX XX
XX DE Pea pra2 light-repressible promoter clone PL4C PCR primer, SEQ ID NO:21.
XX XX
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;

Query Match 100.0%; Score 12; DB 21; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 37 GGATTTTACAGT 26

RESULT 13
AAI97396
ID AAA97396 standard; DNA; 48 BP.
XX AC AAA97396;
XX XX
XX DT 29-JAN-2001 (first entry)
XX XX
XX DE Pea pra2 light-repressible promoter mutant LS4 PCR primer, SEQ ID NO:14.
XX XX
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; mutagenic PCR primer; ss.
XX OS Pisum sativum.
XX OS Synthetic.
XX XX
XX PN WO200055313-A1.
XX XX
XX PD 21-SEP-2000.
XX XX
XX PF 03-MAR-2000; 2000WO-JP01269.
XX XX
XX PR 12-MAR-1999; 99JP-0066551.

```

```

KW deterioration prevention; storage; PCR primer; ss.
XX Pisum sativum.
XX OS
XX PN WO200055313-A1.
XX XX
XX PD 21-SEP-2000.
XX XX
XX PF 03-MAR-2000; 2000WO-JP01269.
XX XX
XX PR 12-MAR-1999; 99JP-0066551.
XX XX
XX PA (SUNR ) SUNTORY LTD.
XX PI Sasaki Y, Nagano Y, Inaba T;
XX XX
XX DR WPI; 2000-587526/55.
XX XX
XX PT New DNA fragment or promoter for expressing a target gene, specifically
XX PT under photoinhibitory conditions, and for transforming a plant cell or
XX PT plant to improve quality and prevent deterioration during storage
XX XX
XX PS Example 3; Page 14; 49pp; Japanese.
XX XX
XX CC The invention relates to a light-repressible promoter (AAA97385), or
XX CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX CC the pea GTP-binding protein pra2. The invention also relates to an
XX CC expression cassette containing the pra2 promoter or its active
XX CC fragments for the expression of a gene under photoinhibitory or dark
XX CC conditions in a plant, and to transgenic plants, their descendants
XX CC and plant tissues comprising the expression cassette. The expression
XX CC cassette of the invention can be used to generate transgenic plants in
XX CC which deterioration during storage in the dark is prevented. This is
XX CC particularly useful for agricultural products. Sequences AAA97387-A97392
XX CC and AAA97398-A97410 represent PCR primers used in an exemplification of
XX CC the invention to amplify the pea pra2 promoter for cloning.
XX XX
XX SQ Sequence 39 BP; 7 A; 8 C; 7 G; 17 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
DB 37 GGATTTTACAGT 26

RESULT 13
AAI97396
ID AAA97396 standard; DNA; 48 BP.
XX AC AAA97396;
XX XX
XX DT 29-JAN-2001 (first entry)
XX XX
XX DE Pea pra2 light-repressible promoter mutant LS4 PCR primer, SEQ ID NO:14.
XX XX
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; mutagenic PCR primer; ss.
XX OS Pisum sativum.
XX OS Synthetic.
XX XX
XX PN WO200055313-A1.
XX XX
XX PD 21-SEP-2000.
XX XX
XX PF 03-MAR-2000; 2000WO-JP01269.
XX XX
XX PR 12-MAR-1999; 99JP-0066551.

```

```

XX PA (SUNR ) SUNTORY LTD.
XX PI
XX PI Sasaki Y, Nagano Y, Inaba T;
XX XX
XX DR WPI; 2000-587526/55.
XX XX
XX PT New DNA fragment or promoter for expressing a target gene, specifically
XX PT under photoinhibitory conditions, and for transforming a plant cell or
XX PT plant to improve quality and prevent deterioration during storage -
XX XX
XX PS Example 3; Page 13; 49pp; Japanese.
XX XX
XX CC The invention relates to a light-repressible promoter (AAA97385), or
XX CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX CC the pea GTP-binding protein pra2. The invention also relates to an
XX CC expression cassette containing the pra2 promoter or its active
XX CC fragments for the expression of a gene under photoinhibitory or dark
XX CC conditions in a plant, and to transgenic plants, their descendants
XX CC and plant tissues comprising the expression cassette. The expression
XX CC cassette of the invention can be used to generate transgenic plants in
XX CC which deterioration during storage in the dark is prevented. This is
XX CC particularly useful for agricultural products. Sequences AAA97393-A97398
XX CC represent PCR primers used in an exemplification of the invention
XX CC to mutate the pea pra2 promoter.
XX SQ Sequence 52 BP; 19 A; 9 C; 12 G; 12 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 52;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db | | | | | | | | | |
13 ggattttacagt 24

RESULT 15
AAA97418/c
ID AAA97418 standard; DNA; 53 BP.
XX AC AAA97418;
XX DT 29-JAN-2001 (first entry)
XX XX
XX DE Pea wild-type pra2 gene light-repressible promoter oligonucleotide, WT4.
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; ss.
XX OS Pisum sativum.
XX XX WO200055313-A1.
XX PN 21-SEP-2000.
XX PD
XX PF 03-MAR-2000; 2000WO-JP01269.
XX XX
XX PR 12-MAR-1999; 99JP-0066551.
XX XX
XX PA (SUNR ) SUNTORY LTD.
XX XX
XX PI Sasaki Y, Nagano Y, Inaba T;
XX DR WPI; 2000-587526/55.
XX XX
XX PT New DNA fragment or promoter for expressing a target gene, specifically
XX PT under photoinhibitory conditions, and for transforming a plant cell or
XX PT plant to improve quality and prevent deterioration during storage -
XX XX
XX PS Example 3; Page 19; 49pp; Japanese.
XX XX
XX CC The invention relates to a light-repressible promoter (AAA97385), or
XX CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
XX CC the pea GTP-binding protein pra2. The invention also relates to an
XX CC expression cassette containing the pra2 promoter or its active
XX CC fragments for the expression of a gene under photoinhibitory or dark
XX CC conditions in a plant, and to transgenic plants, their descendants
XX CC and plant tissues comprising the expression cassette. The expression
XX CC cassette of the invention can be used to generate transgenic plants in
XX CC which deterioration during storage in the dark is prevented. This is
XX CC particularly useful for agricultural products. Sequences AAA97393-A97398
XX CC represent PCR primers used in an exemplification of the invention
XX CC to mutate the pea pra2 promoter.
XX SQ Sequence 48 BP; 15 A; 9 C; 13 G; 11 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 48;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
Db | | | | | | | | | |
13 ggattttacagt 24

RESULT 14
AAA97397
ID AAA97397 standard; DNA; 52 BP.
XX AG AAA97397;
XX XX
XX DT 29-JAN-2001 (first entry)
XX XX
XX DE Pea pra2 light-repressible promoter mutant LS5 PCR primer, SEQ ID NO:15.
XX XX
XX KW GTP-binding protein pra2; pea; light-repressible promoter;
XX KW photoinhibitory; expression cassette; transgenic plant;
XX KW deterioration prevention; storage; mutagenic PCR primer; ss.
XX XX
XX OS Pisum sativum.
XX OS Synthetic.
XX XX WO200055313-A1.
XX PN 21-SEP-2000.
XX PD
XX XX
XX PF 03-MAR-2000; 2000WO-JP01269.
XX XX
XX PR 12-MAR-1999; 99JP-0066551.
XX XX
XX PA (SUNR ) SUNTORY LTD.
XX XX
XX PI Sasaki Y, Nagano Y, Inaba T;
XX DR WPI; 2000-587526/55.
XX XX
XX PT New DNA fragment or promoter for expressing a target gene, specifically
XX PT under photoinhibitory conditions, and for transforming a plant cell or
XX PT plant to improve quality and prevent deterioration during storage -
XX XX

```

CC represent oligonucleotides used in an exemplification of the invention
CC to generate a wild-type pea pr2 promoter fragment.

XX
SQ Sequence 53 BP; 21 A; 11 C; 3 G; 18 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 53;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12
|||||
Db 33 GGATTTTACAGT 22

Search completed: April 3, 2002, 05:18:30
Job time: 2900 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:13:00 ; Search time 1999.71 Seconds
(without alignments)
767.230 Million cell updates/sec

Title: US-09-700-187-2
Perfect score: 93
Sequence: 1 aaagaacacatatattga.....ttacagtaataaagaacga 93

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.om.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.sy.*
28: em.un.*
29: em.vi.*
30: em.htgo_hum.*
31: em.htgo_inv.*
32: em.htgo_rod.*
33: em.htg_hum.*
34: em.htg_inv.*
35: em.htg_rod.*
36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	93	100.0	3441	8	AB007911	AB007911 Pisum sat
2	37.6	40.4	157381	9	AL360267	AL360267 Human DNA
3	37.2	40.0	2498	1	ETHRNA	U03552 Bacillus th
4	36.4	39.1	340	3	AFU77847	U77847 Aconophora
5	36.4	39.1	133021	9	AL161624	AL161624 Human DNA
6	36.4	39.1	169671	2	AC021098	AC021098 Homo sapi
7	35.6	38.3	141279	2	AC023824	AC023824 Homo sapi
8	35.6	38.3	142654	9	AC012049	AC012049 Homo sapi
9	35.6	38.3	154937	9	AC022537	AC022537 Homo sapi
10	35.4	38.1	50143	2	AC027071	AC027071 Homo sapi
11	35.4	38.1	175691	2	AC007615	AC007615 Homo sapi
12	35.4	38.1	191496	2	AC009130	AC009130 Homo sapi
13	35.4	38.1	210043	2	AC009093	AC009093 Homo sapi
14	35	37.6	721	8	AF072528	AF072528 Lycopersi
15	35	37.6	35836	9	AL160414	AL160414 Human DNA
16	34.6	37.2	646	8	AF365230	AF365230 Tetraberl
17	34.6	37.2	66084	8	AB015468	AB015468 Arabidops
18	34.4	37.0	148249	2	AL451133	AL451133 Homo sapi
19	34.4	37.0	157604	9	AC021093	AC021093 Homo sapi
20	34.4	37.0	189938	2	AL355588	AL355588 Homo sapi
21	34.4	37.0	193729	2	AC023772	AC023772 Homo sapi
22	34.2	36.8	293181	1	AP001119	AP001119 Buchnera
23	34	36.6	43152	3	CELF59B1	AF067943 Caenorhab
24	34	36.6	84194	8	AB006701	AB006701 Arabidops
25	34	36.6	104992	2	AC005504	AC005504 Plasmodiu
26	34	36.6	130513	9	AC004592	AC004592 Homo sapi
27	34	36.6	136060	2	AC004153	AC004153 Plasmodiu
28	34	36.6	169546	2	AC004157	AC004157 Plasmodiu
29	34	36.6	245805	2	AC006752	AC006752 Caenorhab
30	34	36.6	300193	3	AE003426	AE003426 Drosophil
31	33.8	36.3	74227	9	HSJ493H23	AL121789 Human DNA
32	33.8	36.3	148018	2	AP003974	AP003974 Oryza sat
33	33.8	36.3	184604	9	AL138682	AL138682 Human DNA
34	33.8	36.3	174617	2	AP003977	AP003977 Oryza sat
35	33.6	36.1	143549	2	AP003910	AP003910 Oryza sat
36	33.6	36.1	168030	2	AC091462	AC091462 Mus muscu
37	33.2	35.7	37219	9	AC091515	AC091515 Homo sapi
38	33.2	35.7	188296	2	AC022324	AC022324 Homo sapi
39	33.2	35.7	188470	2	AC093005	AC093005 Homo sapi
40	33	35.5	656	8	AF365229	AF365229 Tetraberl
41	33	35.5	3707	2	AC012967	AC012967 Drosophil
42	33	35.5	58407	1	MIICG	L71118 Methanococc
43	33	35.5	135599	8	CPU30821	U30821 Cyanophora
44	33	35.5	139064	2	AC073440	AC073440 Homo sapi
45	33	35.5	153549	2	AC022664	AC022664 Homo sapi

ALIGNMENTS

RESULT 1
AB007911
LOCUS Pisum sativum gene for PRA2, complete cds.
DEFINITION Pisum sativum gene for PRA2, complete cds.
ACCESSION AB007911
VERSION AB007911.1 GI:5926717
KEYWORDS PRA2.
SOURCE Pisum sativum DNA.
ORGANISM Pisum sativum

28-SEP-1999

REFERENCE 1 (sites)
AUTHORS Inaba,T., Nagano,Y., Sakakibara,T. and Sasaki,Y.
TITLE Identification of a cis-regulatory element involved in phytochrome down-regulated expression of the pea small GTPase gene praz
JOURNAL Plant Physiol. 120 (2), 491-500 (1999)
MEDLINE 99292944
REFERENCE 2 (bases 1 to 3441)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.

AUTHORS Nagano,Y.
 TITLE Direct Submission
 JOURNAL Submitted (07-OCT-1997) Yukio Nagano, Nagoya University, Graduate School of Bioagricultural Sciences: Chikusa, Nagoya, Aichi 464-8601, Japan (E-mail:nagano@agr.nagoya-u.ac.jp, Tel:81-52-789-4168, Fax:81-52-789-4296)

FEATURES
 source
 1..3441
 /organism="Pisum sativum"
 /db_xref="taxon:3888"
 repeat_unit 904..1018
 gene Join(2326..2567,2729..3188)
 /gene="pra2"
 CDS Join(2326..2567,2729..3188)
 /gene="pra2"
 /note="light-repressible GTP binding protein"
 /codon_start=1
 /product="PRA2"
 /protein_id="BAA84640.1"
 /db_xref="GI:5926718"
 /translation="MNQMGNGVEAEKLOEKIDYVFKVVIGDSAVGKTOILSRFTKNE
 FCFSKSTIGVEFOFKVTYINGKLKAOIWDTAGOERYAVTSAYVRCALGAMLVYDI
 TKROTFDHVAWVEELSHADGSIIVMLIGNKGLDVGQVGTEDAVEFAEDQGLFFS
 ETSAPSGENVSAFLKLLQEKINVKVSRSLCNGKNGKNGHNDHVAALAGEKIDIIISAS
 ELEISEIKKLHSCSC"

BASE COUNT 1143 a 606 c 540 g 1152 t
 ORIGIN

Query Match 100.0%; Score 93; DB 8; Length 3441;
 Best Local Similarity 100.0%; Pred. No. 1.3e-10;
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacatatattgataaattattactaaactattttctagctactgttaac 60
 |||||
 Db 1396 AAAGTAACACATATTTTGATTAATTTATTACTAAACTATTTCTAGTACTGTTAATC 1455
 |||||

QY 61 atgtctgaggattttacagtaataaagaacga 93
 |||||
 Db 1456 ATGTCTGAGGATTTTACAGTAATAAAGAACGA 1488
 |||||

RESULT 2
 AL360267/c 157381 bp DNA PRI 22-MAY-2001
 LOCUS Human DNA sequence from clone RP11-342C20 on chromosome 13,
 DEFINITION complete sequence.
 ACCESSION AL360267
 VERSION AL360267.10 GI:13751441
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 157381)
 Tracey,A.
 Direct Submission
 Submitted (22-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Apr 21, 2001 this sequence version replaced gi:13443424.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
 RP11-342C20 is from the library RPI-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-342C20 The true left end of clone RP11-365J7 is at 99075 in this sequence.

FEATURES
 source
 1..157381
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-342C20"
 /clone_lib="RPI-11.2"
 misc_feature 64251..64259
 /note="Sequence from uni-directional dGTP big dye terminator reads only."
 121811..121816
 /note="Sequence from overlapping clone BA365J7 (AL359472). Assembly confirmed by restriction digest."
 BASE COUNT 51039 a 27179 c 27761 g 51402 t
 ORIGIN

Query Match 40.4%; Score 37.6; DB 9; Length 157381;
 Best Local Similarity 63.0%; Pred. No. 27;
 Matches 58; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 2 aaagtaacacatatattgataaattattactaaactattttctagctactgttaac 61
 |||||
 Db 42229 AAATAACAGACATTTTGATCATTTGTAATAATAGCTAGATAGTAAATGCTACAAA 42170
 |||||

QY 62 tgctcgaggattttacagtaataaagaacga 93
 |||||
 Db 42169 TATTTGAAATCTTTTCATTAGTCAAGCAACA 42138
 |||||

RESULT 3
 BTHKNA 2498 bp DNA BCT 01-APR-1995
 LOCUS Bacillus thuringiensis histidine protein kinase (hknA) gene,
 DEFINITION complete cds.
 ACCESSION U03552
 VERSION U03552.1 GI:495313
 KEYWORDS Bacillus thuringiensis.
 SOURCE Bacillus thuringiensis.
 ORGANISM Bacillus thuringiensis
 Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
 1 (bases 1 to 2498)
 Baum,J.A.
 Tn3401, a new class II transposable element from Bacillus
 thuringiensis
 J. Bacteriol. 176 (10), 2835-2845 (1994)
 MEDLINE 94245608
 REFERENCE 2 (bases 1 to 2498)
 AUTHORs Baum,J.A.
 TITLE Direct Submission
 JOURNAL Submitted (17-NOV-1993) Jim A. Baum, Ecogen Inc., Strain
 Development, Molecular Genetics, 2005 Cabot Boulevard West,
 Langhorne, PA 19047-1810, USA
 LOCATION/Qualifiers
 1..2498
 /organism="Bacillus thuringiensis"

gene /strain="EG7566"
/sub_species="kurstaki"
/db_xref="taxon:1428"

682..1800

CDS /gene="hkna"

682..1800

/gene="hkna"

/standard_name="Hkna"

/codon_start=1

/transl_table=11

/function="phosphorylation"

/evidence-experimental

/product="histidine protein kinase"

/protein_id="AAA64586.1"

/db_xref="GI:520402"

/translation="MEVFPDKDKIKFCSHLKNNRHOFVNNKMKMIISEKDPFKLE
VWONGDLLELLELMEDKDINVLQPCIEKIALERAGADANIGDFVYNNVGRNELF
EAMCELDVSARELKPINAKIHTCFDKLIYTVLKYSILSKNLEEKQOYINETHKRL
TILGMSASVHEFRNPLTSIMGFVKLLKADHPSLSYLDLIISHELDQLNKRSIOFLV
SKEMWNESEFRWLNDFODIIQFLYPSLVNANVLIENKLPYIPLYGVSEYRQVEL
NILMNSIDALESMKEERKIIIDVFEEQDAIRIVIKNGPMIPAENVETIPEPVTTKK
LSTGIGLVCKQIVKEHNGSIMCRSDNDWTEFQIAFOK"

BASE COUNT

ORIGIN

Query Match 40.0%; Score 37.2; DB 1; Length 2498;
Best Local Similarity 63.3%; Pred. No. 56;
Matches 57; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1 aaagtaacacatatttttgataaaattttactaaactattttctagctactgttaac 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1042 AAATACATACCTGTTTGACAAATTAATTATTATTATACCGTTTTAAATACTCGGAATT 1101
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 atgtctgaggattttacagtaataaagaa 90
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1102 ATATCGAAGAATTTAGAGGAAAAACAGCA 1131

RESULT 4

AFU77847 AFU77847 340 bp DNA INV 23-DEC-1996

LOCUS Aconophora ferruginea 12S mitochondrial ribosomal RNA, small

DEFINITION subunit, mitochondrial gene, partial sequence.

ACCESSION U77847

VERSION U77847.1

KEYWORDS GI:1750156

SOURCE Aconophora ferruginea.

ORGANISM Mitochondrion Aconophora ferruginea

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Paraneoptera; Hemiptera; Euhemiptera;

Membracidae; Aconophora.

1 (bases 1 to 340)

Liu, D.

Direct Submission

TITLE Submitted (12-NOV-1996) Exploratory Research, Human Genome

JOURNAL Sciences, 9410 Key West Ave., Rockville, MD 20850, USA

FEATURES Location/Qualifiers

source 1..340

/organism="Aconophora ferruginea"

/organella="mitochondrion"

/db_xref="taxon:54606"

<1..>340

/product="12S small subunit ribosomal RNA"

BASE COUNT 131 a 25 c 43 g 141 t

ORIGIN

Query Match 39.1%; Score 36.4; DB 3; Length 340;

Best Local Similarity 66.7%; Pred. No. 1.1e+02;

Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 12 atatttgataaattttactaaactattttctagctactgttaacatctgtctgagga 71

Db 228 AAATATGGTTTACTTTTATTAGTAATAATTATTAACCTAGATGTTTAAATTTAGGA 287
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 72 ttttacagtaataaagaa 89
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 288 TTTAAAGCTAATAATAAAA 305

RESULT 5

LOCUS AL161624/c

DEFINITION Human DNA sequence from clone RP11-48719 on chromosome X, complete

sequence.

AL161624

ACCESSION AL161624.7

VERSION GI:14160916

KEYWORDS HTG.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 133021)

AUTHORS Bird, C.

TITLE Direct Submission

JOURNAL Submitted (18-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On May 20, 2001 this sequence version replaced gi:13274323.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

This sequence was generated from part of bacterial clone contigs of human

chromosome X, constructed by the Sanger Centre Chromosome X Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX

RP11-48719 is from the library RPCI-11.2 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone

RP11-48719. It may be shorter because we sequence overlapping

sections only once, except for a 100 base overlap.

The true left end of clone RP11-212B22 is at 132922 in this

sequence. The true right end of clone RP11-274M8 is at 100 in this

sequence.

FEATURES Location/Qualifiers

source 1..133021

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone="RP11-48719"

/clone_lib="RPCI-11.2"

118..425

/note="Alusq repeat: matches 1..300 of consensus"

982..1042

/note="MER69 repeat: matches 1..62 of consensus"

1311..1622

/note="AlusSp repeat: matches 1..310 of consensus"

2450..2552

repeat_region	/note="L2 repeat: matches 2636. .2750 of consensus" 5104. .5413
repeat_region	/note="AluSp repeat: matches 1. .304 of consensus" 5900. .6018
repeat_region	/note="L2 repeat: matches 2486. .2568 of consensus" 6019. .6316
repeat_region	/note="AluSp repeat: matches 1. .304 of consensus" 6317. .6514
repeat_region	/note="L2 repeat: matches 2187. .2486 of consensus" 7589. .7901
repeat_region	/note="AluSq repeat: matches 1. .313 of consensus" 8993. .9101
repeat_region	/note="MIR repeat: matches 156. .262 of consensus" 9102. .9441
repeat_region	/note="Tigger3b repeat: matches 887. .1231 of consensus" 9445. .9971
repeat_region	/note="Tigger3b repeat: matches 2. .540 of consensus" 9974. .10116
repeat_region	/note="MIR repeat: matches 1. .141 of consensus" 10516. .10812
repeat_region	/note="AluX repeat: matches 1. .297 of consensus" 10815. .11134
repeat_region	/note="AluJo repeat: matches 1. .291 of consensus" 11810. .12051
repeat_region	/note="MIR repeat: matches 23. .262 of consensus" 13800. .14023
repeat_region	/note="MER53 repeat: matches 1. .189 of consensus" 14038. .14282
repeat_region	/note="MIR repeat: matches 10. .262 of consensus" 14661. .14723
repeat_region	/note="L1MB7 repeat: matches 6105. .6169 of consensus" 14724. .14834
repeat_region	/note="AluSq/x repeat: matches 1. .111 of consensus" 14836. .14980
repeat_region	/note="MER20 repeat: matches 59. .218 of consensus" 15747. .15910
repeat_region	/note="MIR repeat: matches 84. .255 of consensus" 17000. .17310
repeat_region	/note="AluX repeat: matches 1. .312 of consensus" 17611. .17646
repeat_region	/note="MIR repeat: matches 199. .234 of consensus" 17647. .17944
repeat_region	/note="AluSq repeat: matches 2. .299 of consensus" 17945. .18142
repeat_region	/note="MIR repeat: matches 1. .199 of consensus" 18160. .18382
repeat_region	/note="MIR repeat: matches 30. .262 of consensus" 18803. .19103
repeat_region	/note="AluSq repeat: matches 1. .301 of consensus" 19223. .19345
repeat_region	/note="L1PA7 repeat: matches 6010. .6141 of consensus" 19470. .19565
repeat_region	/note="AluSq/x repeat: matches 212. .307 of consensus" 19571. .19723
repeat_region	/note="L2 repeat: matches 2008. .2143 of consensus" 19724. .20154
repeat_region	/note="MLT2FB repeat: matches 3. .414 of consensus" 20155. .20431
repeat_region	/note="L2 repeat: matches 1666. .2008 of consensus" 20458. .20767
repeat_region	/note="MLT1C repeat: matches 131. .466 of consensus" 20768. .20931
repeat_region	/note="FRAM repeat: matches 0. .163 of consensus" 21050. .21545
repeat_region	/note="Charlielb repeat: matches 14. .523 of consensus" 21602. .21901
repeat_region	/note="AluSq1 repeat: matches 1. .300 of consensus" 22056. .22620
repeat_region	/note="LTR26 repeat: matches 43. .603 of consensus" 22624. .22930
repeat_region	/note="Aluwb repeat: matches 1. .311 of consensus" 22938. .23111
repeat_region	/note="AluSq/x repeat: matches 137. .310 of consensus" 22938. .23111

repeat_region	23112..23183 /note="LRR26 repeat: matches 1. .72 of consensus" 24535..24701 /note="MTR repeat: matches 57. .232 of consensus" 25142..25235 /note="MIR repeat: matches 47. .140 of consensus" 27313..27526 /note="MTR repeat: matches 31. .261 of consensus" 29016..29069 /note="AluSg repeat: matches 247. .300 of consensus" 31157..31458 /note="AluSg1 repeat: matches 1. .297 of consensus" 32799..33196 /note="LIM49 repeat: matches 5892. .6308 of consensus" 33312..33444 /note="FLAM_A repeat: matches 1. .133 of consensus" 34016..34172 /note="MER5A repeat: matches 32. .185 of consensus" 34623..34658 /note="9 copies 4 mer tgt 91% conserved" 35365..35535 /note="MR45 repeat: matches 1. .171 of consensus" 36270..36567 /note="AluSp repeat: matches 1. .298 of consensus" 41820..46785 /note="LIP46 repeat: matches 11. .4967 of consensus" 46786..47086 /note="AluY repeat: matches 3. .299 of consensus" 47087..48267 /note="LIP46 repeat: matches 4967. .6141 of consensus" 48774..48803 /note="LIM4c repeat: matches 905. .934 of consensus" 49784..49831 /note="12 copies 4 mer caca 93% conserved" 49863..50116 /note="AluXs repeat: matches 57. .309 of consensus" 50156..50344 /note="MR44 repeat: matches 1. .183 of consensus" 50342..50612 /note="MR44 repeat: matches 421. .728 of consensus" 51509..51638 /note="FLAM_A repeat: matches 1. .127 of consensus" 51679..51898 /note="MR58B repeat: matches 114. .341 of consensus" 52004..52183 /note="MIR repeat: matches 50. .262 of consensus" 54482..54616 /note="LIME2 repeat: matches 6011. .6145 of consensus" 55248..55774 /note="LRR40a repeat: matches 8. .517 of consensus" 57036..57339 /note="AluJB repeat: matches 1. .294 of consensus" 57676..57908 /note="LIM48 repeat: matches 5986. .6242 of consensus" 57912..58211 /note="AluXs repeat: matches 5. .298 of consensus" 58220..58259 /note="20 copies 2 mer tg 90% conserved" 58373..58500 /note="L2 repeat: matches 1651. .1778 of consensus" 58705..59004 /note="AluXs repeat: matches 1. .295 of consensus" 59121..59259 /note="MR33 repeat: matches 7. .149 of consensus"
---------------	---

```

Query Match      39.1%; Score 36.4; DB 9; Length 133021;
Best Local Similarity 66.7%; Pred. No. 49;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 4 agtaacacatatattgataaaattattactaaaactattttctagtaactgttaatactg 63
      ||| |||||
Db 77449 AGAAACACATGATCTGGAACTTCATTTCTAAATTTATTTTTTCTAAATTTCTAAATTT 77390

```

```

QY      64  ttgaggattttacagta 81
      || | | || | | |
Db 77389 ACTCACTAATTTATTATA 77372

RESULT 6
AC021098/c
LOCUS   AC021098.3 GI:7230836
DEFINITION Homo sapiens chromosome X clone RP11-33A2, WORKING DRAFT SEQUENCE,
18 unordered pieces.
ACCESSION AC021098
VERSION   AC021098.3
KEYWORDS  HTG; HTGS_PRAISE1; HTGS_DRAFT.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169671)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 169671)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:6922906.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0033A02
----- Summary Statistics -----
Sequencing vector: M13; 81%
Sequencing vector: plasmid; 19%
Chemistry: Dye-primer ET; 81% of reads
Chemistry: Dye-terminator Big Dye; 19% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 159126 bases at least Q40
Consensus quality: 162659 bases at least Q30
Consensus quality: 164792 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 167971; sum-of-contigs
Quality coverage: 4.16 in Q20 bases; agarose-fp
Quality coverage: 4.21 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1196: contig of 1196 bp in length
* 1197: gap of unknown length
* 1297: contig of 1988 bp in length
* 3285: gap of unknown length
* 3385: contig of 2639 bp in length
* 6024: gap of unknown length
* 6124: contig of 2645 bp in length
* 8769: gap of unknown length
* 8868: contig of 3303 bp in length
* 8869: gap of unknown length
* 12172: contig of 3706 bp in length
* 12272: gap of unknown length
* 15978: contig of 8173 bp in length
* 16078: gap of unknown length
* 24251: contig of 8331 bp in length
* 32681: gap of unknown length
* 32682: gap of unknown length

```

```

* 32782 41939: contig of 9158 bp in length
* 41940 42039: gap of unknown length
* 42040 48802: contig of 6763 bp in length
* 48803 48902: gap of unknown length
* 48903 57427: contig of 8525 bp in length
* 57428 57527: gap of unknown length
* 57528 66323: contig of 8796 bp in length
* 66324 78283: gap of unknown length
* 78284 78384: gap of unknown length
* 78384 93662: contig of 15279 bp in length
* 93663 93762: gap of unknown length
* 93763 111527: contig of 17764 bp in length
* 111527 129715: gap of unknown length
* 129715 129815: contig of 18088 bp in length
* 129815 144538: gap of unknown length
* 144538 14638: contig of 14723 bp in length
* 14638 169671: contig of 25034 bp in length.

FEATURES             Location/Qualifiers
     source            1..169671
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-33A2"
     misc_feature      1..1196
                     /note="assembly_name:Contig7"
     misc_feature      1297..3284
                     /note="assembly_name:Contig8"
     misc_feature      3385..6023
                     /note="assembly_name:Contig9"
     misc_feature      6124..8768
                     /note="assembly_name:Contig10"
     misc_feature      8869..12171
                     /note="assembly_name:Contig11"
     misc_feature      12272..15977
                     /note="assembly_name:Contig12"
     misc_feature      16078..24250
                     /note="assembly_name:Contig13"
     misc_feature      24351..32681
                     /note="assembly_name:Contig14"
     misc_feature      32782..41939
                     /note="assembly_name:Contig15"
     misc_feature      42040..48802
                     /note="assembly_name:Contig16"
                     clone_end:Sp6
                     vector_side:right
     misc_feature      48903..57427
                     /note="assembly_name:Contig17"
                     clone_end:T7
                     vector_side:right
     misc_feature      57528..66323
                     /note="assembly_name:Contig18"
     misc_feature      66424..78283
                     /note="assembly_name:Contig19"
     misc_feature      78384..93662
                     /note="assembly_name:Contig20"
     misc_feature      93763..111526
                     /note="assembly_name:Contig21"
     misc_feature      111627..129714
                     /note="assembly_name:Contig22"
     misc_feature      129815..144537
                     /note="assembly_name:Contig23"
     misc_feature      144638..169671
                     /note="assembly_name:Contig24"

BASE COUNT  53919 a 31234 c 31915 g 50898 t 1705 others
ORIGIN

```

```

Query Match          39.1%; Score 36.4; DB 2; Length 169671;
Best Local Similarity 66.7%; Pred. NO. 48;
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```



```

BASE COUNT      51071 a 29071 c 26776 g 48019 t
ORIGIN

Query Match      38.3%; Score 35.6; DB 9; Length 154937;
Best Local Similarity 62.2%; Pred. No. 71;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 aaagtaacacataatttggatataatttattactaaaactattttctagctactgttaac 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87539 AAAAAATTAACCCCATGTTACTTTTTCAGTAGTACTAAAAAAGTATTTTAAATTTATGTGCAATC 87598

QY 61 atgtctgagagattttacagtaataaagaaa 90
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87599 ATGTTAGTAACTACTAAAAACAAGAAAGAAA 87628

RESULT 10
AC027071/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC027071 50143 bp DNA HTG 27-MAR-2000
Homo sapiens chromosome 9 clone RP11-200D20 map 9, LOW-PASS
SEQUENCE SAMPLING.
AC027071
AC027071.1 GI:7330331
HTG; HTGS_PHASE0.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 50143)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 9, clone RP11-200D20
2 (bases 1 to 50143)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
Boguski,M., Bouckgeer,B., Brown,A., Burkett,G.,
Campione,A., Castle,A., Choehel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArnell,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.P. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7662
Center clone name: 200_D_20
-----
* NOTE: This record contains 57 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads

```



```
* 14646 14745: gap of unknown length
* 14751: contig of 1006 bp in length
* 15752 15851: gap of unknown length
* 15852 16968: contig of 1017 bp in length
* 16969 16968: gap of unknown length
* 19037: contig of 2069 bp in length
* 19038 19137: gap of unknown length
* 19138 20794: contig of 1657 bp in length
* 20795 20894: gap of unknown length
* 20895 22876: contig of 1982 bp in length
* 22877 22976: gap of unknown length
* 22977 24751: contig of 1775 bp in length
* 24752 24851: gap of unknown length
* 24852 27381: contig of 2530 bp in length
* 27382 27481: gap of unknown length
* 27482 30339: contig of 2858 bp in length
* 30340 30439: gap of unknown length
* 30440 32397: contig of 1958 bp in length
* 32398 32497: gap of unknown length
* 32498 34880: contig of 2383 bp in length
* 34881 34980: gap of unknown length
* 34981 37578: contig of 2598 bp in length
* 37579 37678: gap of unknown length
* 37679 40004: contig of 2326 bp in length
* 40005 40104: gap of unknown length
* 40105 43544: contig of 3440 bp in length
* 43545 43644: gap of unknown length
* 43645 46429: contig of 2785 bp in length
* 46430 46529: gap of unknown length
* 46530 49886: contig of 3157 bp in length
* 49887 49786: gap of unknown length
* 49787 52269: contig of 2483 bp in length
* 52270 52369: gap of unknown length
* 52370 55195: contig of 2826 bp in length
* 55196 55295: gap of unknown length
* 55296 58116: contig of 3221 bp in length
* 58117 58616: gap of unknown length
* 58617 61359: contig of 2743 bp in length
* 61360 61459: gap of unknown length
* 61460 66151: contig of 4692 bp in length
* 66152 66251: gap of unknown length
* 66252 73145: contig of 6894 bp in length
* 73146 73245: gap of unknown length
* 73246 78200: contig of 4955 bp in length
* 78201 83186: contig of 4886 bp in length
* 83187 83286: gap of unknown length
* 83287 88570: contig of 5284 bp in length
* 88571 88671: gap of unknown length
* 88672 95276: contig of 6606 bp in length
* 95277 95376: gap of unknown length
* 95377 102234: contig of 6858 bp in length
* 102235 102334: gap of unknown length
* 102335 107630: contig of 5296 bp in length
* 107631 107730: gap of unknown length
* 107731 120220: contig of 12490 bp in length
* 120221 120320: gap of unknown length
* 120321 134451: contig of 14131 bp in length
* 134452 134551: gap of unknown length
* 134552 144891: contig of 10340 bp in length
* 144892 144991: gap of unknown length
* 144992 159434: contig of 14443 bp in length
* 159435 159534: gap of unknown length
* 159535 175863: contig of 16329 bp in length
* 175864 175963: gap of unknown length
* 175964 191496: contig of 15533 bp in length.
Location/Qualifiers
1. .191496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-501P17"
/clone_lib="RPC1 human BAC library 11"
```

FEATURES
source

```
BASE COUNT 47031 a 47510 c 47005 g 45732 t 4218 others
ORIGIN
Query Match 38.1%; Score 35.4; DB 2; Length 191496;
Best Local Similarity 66.2%; Pred. No. 76;
Matches 51; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 2 aaagtaacacatatcttgataaattattactaaactatttcttagtacttgtaatca 61
||||| ||||||| ||||||| ||||||| ||||||| |||
Db 161383 ATAGTAGCTCATATTTGTTAAATTTATCCCTAAGAATTTATGATTCATTGCAATGG 161324
||||| ||||||| ||||||| ||||||| ||||||| |||
QY 62 tgtctgagagatttaca 78
||||| ||||| ||
Db 161323 TATTTTATTTTGTGCA 161307

RESULT 13
AC009093/c
LOCUS AC009093 210043 bp DNA HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-426C22, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
ACCESSION AC009093
VERSION AC009093.7 GI:13786300
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 210043)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 210043)
REFERENCE DOE Joint Genome Institute.
DIRECT SUBMISSION
SUBMITTED (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
ON Apr 25, 2001 this sequence version replaced gi:9954601.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 570019
Center clone name: RPC1-11_426C22
-----
Summary Statistics
Consensus quality: 193032 bases at least Q40
Consensus quality: 199717 bases at least Q30
Consensus quality: 202073 bases at least Q20
Estimated insert size: 239000; agarose-fp estimation
Estimated insert size: 208543; sum-of-contigs estimation
Quality coverage: 5.83 in Q20 bases; agarose-fp estimation
Quality coverage: 6.68 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1678: contig of 1678 bp in length
* 1679 1778: gap of unknown length
* 1779 3884: contig of 2106 bp in length
* 3885 3985: gap of unknown length
* 3985 5471: contig of 1487 bp in length
* 5472 5571: gap of unknown length
* 5572 6652: contig of 1081 bp in length
* 6653 6752: gap of unknown length
* 6753 8717: contig of 1965 bp in length
```


* 8718 8817: gap of unknown length
 * 8818 10104: contig of 1287 bp in length
 * 10105 10204: gap of unknown length
 * 10205 12294: contig of 2090 bp in length
 * 12295 12394: gap of unknown length
 * 12395 15488: contig of 3094 bp in length
 * 15489 15588: gap of unknown length
 * 15589 18845: contig of 3257 bp in length
 * 18846 18945: gap of unknown length
 * 18946 23168: contig of 4223 bp in length
 * 23169 23268: gap of unknown length
 * 23269 29524: contig of 6256 bp in length
 * 29525 29624: gap of unknown length
 * 29625 54542: contig of 24918 bp in length
 * 54543 54642: gap of unknown length
 * 54643 81600: contig of 26958 bp in length
 * 81601 81700: gap of unknown length
 * 81701 115005: contig of 33305 bp in length
 * 115006 115105: gap of unknown length
 * 115106 161572: contig of 46467 bp in length
 * 161573 161672: gap of unknown length
 * 161673 210043: contig of 48371 bp in length.

FEATURES

source

1. .210043
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-426C22"
 /clone_lib="RPC1 human BAC library 11"
 53123 a 52905 c 51736 g 50777 t 1502 others

BASE COUNT

ORIGIN

Query Match 38.1%; Score 35.4; DB 2; Length 210043;
 Best Local Similarity 66.2%; Pred. No. 75;
 Matches 51; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 2 aaagtaacacatttttgataaaattttattactaaacattttctagctactgttaataca 61
 Db 110362 ATAGTACCTATATTTTGTAAATTTATCCCTAAGAAATTTATGATTCACCTGCAATGG 110303
 QY 62 tgtctgaggattttaca 78
 Db 110302 TATTATTATTTTGTGCA 110286

RESULT 14

AF072528

LOCUS Lycopersicon pennellii clone AG15 paracentromeric sequence.
 DEFINITION Lycopersicon pennellii clone AG15 paracentromeric sequence.
 ACCESSION AF072528
 VERSION AF072528.1 GI:3342056
 KEYWORDS Lycopersicon pennellii.
 SOURCE Lycopersicon pennellii.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

REFERENCE

1 (bases 1 to 721)
 Weide,R., Hontelez,J., van Kammen,A., Koornneef,M. and Zabel,P.
 Paracentromeric sequences on tomato chromosome 6 show homology to
 human satellite III and to the mammalian CENP-B binding box

JOURNAL

MEDLINE

98418482
 2 (bases 1 to 721)

REFERENCE

Weide,R., Hontelez,J., van Kammen,A., Koornneef,M. and Zabel,P.
 Direct Submission
 Submitted (19-JUN-1998) Phytopathology, Wageningen Agricultural
 University, Binnenhaven 9, Wageningen 6709 PD, The Netherlands

JOURNAL

FEATURES

source

1. .721
 /organism="Lycopersicon pennellii"

/strain="LA 716"
 /db_xref="taxon:28526"
 /chromosome="6"
 /map="centromere"
 /clone="AG15"
 239 a 69 c 116 g 297 t

BASE COUNT

ORIGIN

Query Match 37.6%; Score 35; DB 8; Length 721;
 Best Local Similarity 61.5%; Pred. No. 1.9e+02;
 Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 3 aagtaacacatttttgataaaattttattactaaacattttctagctactgttaataca 62
 Db 326 AATAACCCATAGTTTATTTTATATATTTAAGTAAAGATTGTCCAGTCATTTTAATTTT 385

QY 63 gtctgaggattttacagtaataaagaacga 93

Db 386 TAATATTAGTTTATATATATGATAATGTAACATA 416

RESULT 15

AL160414/c

LOCUS

DEFINITION

AL160414 35836 bp DNA PRI 15-FEB-2001
 Human DNA sequence from clone RP5-931H19 on chromosome 20. Contains
 ESTs, GSSs and CpG islands. Contains the OXT gene for oxytocin
 (neurophysin I) and the AVP gene for arginine vasopressin
 (neurophysin II), complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL160414.18 GI:9863596
 HTG; AVP; CpG island; neurophysin; OXT; oxytocin; vasopressin.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 (bases 1 to 35836)
 Heath,P.
 Direct Submission
 Submitted (13-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 21, 2000 this sequence version replaced gi:9843543.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 IMPORTANT: This sequence is not the entire insert of clone
 RP5-931H19 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP5-1187M17 is at 35737 in this
 sequence. The true right end of clone RP4-534B8 is at 100 in this
 sequence. This sequence was finished as follows unless otherwise
 noted: all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.
 RP5-931H19 is from the library RPC1-5 constructed by the group of
 Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm

FEATURES VECTOR: pCYPAC2.
 Location/Qualifiers
 1..35836
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP5-931H19"
 /clone_lib="RPC1-5"
 1..158
 /note="AluY repeat: matches 1..158 of consensus"
 repeat_region
 2142..2408
 /note="AluX repeat: matches 1..296 of consensus"
 repeat_region
 2834..3122
 /note="AluX repeat: matches 1..290 of consensus"
 repeat_region
 3133..3363
 /note="AluX repeat: matches 1..231 of consensus"
 repeat_region
 3557..3870
 /note="AluY repeat: matches 1..307 of consensus"
 repeat_region
 4187..4417
 /note="MIR repeat: matches 40..262 of consensus"
 repeat_region
 4419..4723
 /note="AluSp repeat: matches 1..312 of consensus"
 repeat_region
 4954..5248
 /note="AluSp repeat: matches 3..298 of consensus"
 repeat_region
 5282..5593
 /note="AluX repeat: matches 1..310 of consensus"
 repeat_region
 5594..5704
 /note="L1MC4 repeat: matches 7701..7813 of consensus"
 repeat_region
 5766..5904
 /note="FLAM_C repeat: matches 1..133 of consensus"
 repeat_region
 5905..6242
 /note="L1MA5A repeat: matches 5478..5827 of consensus"
 repeat_region
 6243..6546
 /note="AluX repeat: matches 18..309 of consensus"
 repeat_region
 6631..6679
 /note="L1M4 repeat: matches 2698..2746 of consensus"
 misc_feature
 6701..7023
 /note="match: GSS: Em:AQ318472"
 misc_feature
 6701..7051
 /note="match: GSS: Em:AQ782226"
 repeat_region
 6721..7033
 /note="AluSp repeat: matches 1..313 of consensus"
 misc_feature
 complement(7032..7218)
 /note="match: GSS: Em:AQ236587 Em:AQ239242"
 misc_feature
 complement(7033..7172)
 /note="match: GSS: Em:AQ423172"
 misc_feature
 complement(7034..7218)
 /note="match: GSS: Em:AQ286166 Em:AQ596537"
 misc_feature
 complement(7037..7219)
 /note="match: GSS: Em:AQ614522"
 misc_feature
 complement(7045..7218)
 /note="match: GSS: Em:AQ545857"
 misc_feature
 7048..7220
 /note="match: GSS: Em:AQ044844"
 repeat_region
 7233..7297
 /note="L1M1 repeat: matches 1190..1252 of consensus"
 misc_feature
 complement(join(7233..7296,7617..7774))
 /note="match: GSS: Em:AQ541198"
 repeat_region
 7298..7614
 /note="AluJo repeat: matches 6..307 of consensus"
 repeat_region
 7615..7798
 /note="L1M1 repeat: matches 998..1190 of consensus"
 repeat_region
 7800..7932
 /note="AluSg/x repeat: matches 1..133 of consensus"
 repeat_region
 7937..8074
 /note="AluSg repeat: matches 1..140 of consensus"
 repeat_region
 8075..8382
 /note="AluVa5 repeat: matches 5..311 of consensus"
 repeat_region
 8383..8550
 /note="AluSg repeat: matches 140..302 of consensus"
 repeat_region
 8730..9040
 /note="AluY repeat: matches 295..538 of consensus"
 repeat_region
 9060..9207
 /note="AluJo repeat: matches 1..305 of consensus"

/note="HERV16 repeat: matches 2908..3069 of consensus"
 9412..9705
 /note="AluSx repeat: matches 16..310 of consensus"
 9717..10018
 /note="AluSg repeat: matches 1..306 of consensus"
 10020..10072
 /note="L1MA9 repeat: matches 6259..6308 of consensus"
 10218..10818
 /note="L1M2 repeat: matches 95..425 of consensus"
 10827..10929
 /note="L1M1 repeat: matches 2202..2684 of consensus"
 10993..11252
 /note="AluSg repeat: matches 1..301 of consensus"
 11320..12087
 /note="L1 repeat: matches 3006..3748 of consensus"
 12191..12246
 /note="L1P5 repeat: matches 4436..4491 of consensus"
 12247..12552
 /note="AluSg repeat: matches 3..306 of consensus"
 12553..13393
 /note="L1P5 repeat: matches 4491..5324 of consensus"
 13394..13696
 /note="AluSg repeat: matches 1..304 of consensus"
 13697..13739
 /note="L1P5 repeat: matches 5324..5366 of consensus"
 13741..13842
 /note="AluJo/FRAM repeat: matches 198..299 of consensus"
 13843..14144
 /note="AluY repeat: matches 1..299 of consensus"
 14149..14687
 /note="SVA repeat: matches 8..548 of consensus"
 14692..14799
 /note="L1P5 repeat: matches 5315..5425 of consensus"
 14815..15106
 /note="AluSx repeat: matches 1..294 of consensus"
 15108..15271
 /note="AluY repeat: matches 133..307 of consensus"
 15272..15296
 /note="L1M1 repeat: matches 5807..5828 of consensus"
 15297..16003
 /note="L1T8 repeat: matches 1..691 of consensus"
 16004..16110
 /note="L1M1 repeat: matches 5828..5952 of consensus"
 16191..16264
 /note="L1M3 repeat: matches 6056..6129 of consensus"
 16324..16396
 /note="L2 repeat: matches 1370..1454 of consensus"
 16436..16508
 /note="L2 repeat: matches 1582..1649 of consensus"
 16509..16811
 /note="AluSp repeat: matches 1..306 of consensus"
 16812..17029
 /note="L2 repeat: matches 1649..1896 of consensus"
 17030..17337
 /note="AluSx repeat: matches 1..307 of consensus"
 17346..17497
 /note="AluSg/x repeat: matches 135..288 of consensus"
 17499..17688
 /note="L2 repeat: matches 1881..2104 of consensus"
 17689..17983
 /note="AluY repeat: matches 3..297 of consensus"
 17984..18023
 /note="L2 repeat: matches 2104..2138 of consensus"
 18126..18218
 /note="L2 repeat: matches 2303..2398 of consensus"
 18346..18628
 /note="AluSx repeat: matches 1..299 of consensus"
 18629..18694
 /note="Alu repeat: matches 246..296 of consensus"
 18944..19131
 /note="L1T1E repeat: matches 295..538 of consensus"
 19132..19439
 /note="AluJo repeat: matches 1..305 of consensus"

```
repeat_region 19440..19473
/note="MLTIE repeat: matches 538..564 of consensus"
repeat_region 19506..19744
/note="L2 repeat: matches 2465..2685 of consensus"
repeat_region 20242..20618

Query Match      37.6%; Score 35; DB 9; Length 35836;
Best Local Similarity 70.1%; Pred. No. 1.le+02;
Matches 47; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 13 tatttgataaattattactaaactattttcttagtactgtgtaataatcatgctcgaggat 72
    ||||| || ||||| ||||| ||||| ||||| || ||||| |||
Db 14715 TATTTTATTATTATTTTAAAAAAATTTT TAGTATTTATTGATTCATTCTTGGGTGT 14656

Qy 73 ttacag 79
    || | |
Db 14655 TTCTCGG 14649
```

Search completed: April 3, 2002, 05:14:06
Job time: 4161 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:14:51 ; Search time 81.95 seconds
(without alignments)
257.016 Million cell updates/sec

Title: US-09-700-187-2
Perfect score: 93
Sequence: 1 aaagtaacacatatattga.....ttacagtaataaagaacga 93

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	29.4	31.6	1984	1	US-07-885-970A-25
C 2	29.4	31.6	1985	1	US-08-298-687A-25
C 3	29.4	31.6	1985	1	US-08-298-829-25
C 4	29.4	31.6	6373	3	US-08-462-728-1
C 5	29.4	31.6	6375	1	US-08-168-917-5
C 6	29.4	31.6	6375	2	US-08-460-510-5
C 7	29.4	31.6	6375	2	US-08-460-490-5
C 8	29.4	31.6	6375	5	PCT-US92-00730-5
C 9	29.4	31.6	6375	5	PCT-US92-00862-5
C 10	28.2	30.3	392	4	US-09-385-982-95
C 11	28.2	30.3	4576	1	US-08-832-883-49
C 12	28.2	30.3	4576	2	US-08-832-877-49
C 13	28.2	30.3	4673	1	US-07-638-431-1
C 14	28.2	30.3	4673	5	PCT-US92-00018-1
C 15	28.2	30.3	6138	4	US-09-067-800-4
C 16	28.2	30.3	6138	4	US-09-349-677-4
C 17	28	28	2290	6	5312912-1
C 18	27.6	29.7	1947	3	US-08-604-991-1
C 19	27.6	29.7	1947	3	US-09-363-639-1
C 20	27.4	29.5	617	4	US-09-385-982-222
C 21	27.4	29.5	967	4	US-08-960-780-47
C 22	27.4	29.5	967	4	US-09-073-898-47
C 23	27.2	29.2	713	4	US-08-998-416-956
C 24	27.2	29.2	3974	4	US-08-467-504-3
C 25	27.2	29.2	5336	4	US-09-102-528-11
C 26	27.2	29.2	6471	4	US-09-353-585-1
C 27	27	29.0	3095	6	5231168-1

C 28	26.8	28.8	1431	4	US-09-316-083-2	Sequence 2, Appli
C 29	26.6	28.6	591	4	US-09-328-111-807	Sequence 807, App
C 30	26.6	28.6	663	4	US-08-998-416-187	Sequence 187, App
C 31	26.6	28.6	696	4	US-08-998-416-779	Sequence 779, App
C 32	26.6	28.6	719	4	US-08-998-416-1138	Sequence 1138, Ap
C 33	26.6	28.6	856	4	US-08-998-416-289	Sequence 289, App
C 34	26.6	28.6	2481	2	US-08-958-642-3	Sequence 3, Appli
C 35	26.6	28.6	2481	3	US-08-778-394-1	Sequence 1, Appli
C 36	26.6	28.6	2481	3	US-08-778-423A-3	Sequence 3, Appli
C 37	26.6	28.6	13158	2	US-08-687-080-105	Sequence 105, App
C 38	26.4	28.4	662	4	US-08-998-416-185	Sequence 185, App
C 39	26.4	28.4	701	4	US-08-998-416-701	Sequence 701, App
C 40	26.4	28.4	724	4	US-08-998-416-683	Sequence 683, App
C 41	26.4	28.4	732	4	US-08-998-416-1036	Sequence 1036, Ap
C 42	26.4	28.4	767	4	US-08-998-416-472	Sequence 472, App
C 43	26.4	28.4	782	4	US-08-998-416-224	Sequence 224, App
C 44	26.4	28.4	827	4	US-08-998-416-535	Sequence 535, App
C 45	26.4	28.4	828	4	US-08-998-416-538	Sequence 538, App

ALIGNMENTS

RESULT 1
US-07-885-970A-25/c
; Sequence 25, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1984 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium barbadense
; STRAIN: Sea Island
; IMMEDIATE SOURCE:
; LIBRARY: EMBL SI

Best Local Similarity	58.6%	Pred. No. 8.1;
Matches	51: Conservative	0: Mismatches
		36: Indels
		0: Gaps
		0: Gaps

```

Qy 1 aaagtaacacatattttgataaaacttattctactaaaaactattttctagtacttgtaa 60
    ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1023 AAAATTAAATTTAATAATTATAAAAACCACTTAATAAAAAATTATATTTAAGTCTGCTG 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 atgtcagggaattttcacagtaataaag 87
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 963 AGGCTTAAATTTCAACCGATAATAAAG 937
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-08-462-728-1
; Sequence 1, Application US/08462728
; Patent No. 6043211
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, Lewis T.
; APPLICANT: ESCOBEDO, Jaime A.
; TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR RECEPTORS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market, Steuart Street Tower, 20th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,728
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/309,322
; FILING DATE: 10-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/151,414
; FILING DATE: 02-FEB-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-267-2-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/326-2400
; TELEFAX: 415/326-2422
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 129..3398
US-08-462-728-1

Query Match 31.6%; Score 29.4; DB 3; Length 6373;
Best Local Similarity 58.6%; Pred.No. 8.9; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 36;

Qy 4 agtaaacacatattttgataaaacttattctactaaaaactattttctagtacttgtaa 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6027 AGTTATACATATATACATAAAGATATATCTGAACCTCTTATGACGGTTTTTGTAATACTG 6086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 64 tctgagggaatttcacagtaataaagaa 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6087 TTCACATAGTAGCGGAAGCAAAATATA 6113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 6089 TTCGACATAGTGACGGAAGCAAATATA 6115

RESULT 6

```

US-08-460-510-5
; Sequence 5, Application US/08460510
; Patent No. 5872218
; GENERAL INFORMATION:
;
; APPLICANT: Wolf, David
; APPLICANT: Tomlinson, James E.
; APPLICANT: Fretto, Larry J.
; APPLICANT: Giese, Neill A.
; APPLICANT: Escobedo, Jaime A.
; APPLICANT: Williams, Lewis T.
;
; TITLE OF INVENTION: DOMAINS OF EXTRACELLULAR REGION OF HUMAN
;
; TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR RECEPTOR POLYPEPTIDES
;
; NUMBER OF SEQUENCES: 23
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW
;
; STREET: One Market Plaza, Steuart Street Tower, Suite 2000
;
; CITY: San Francisco
;
; STATE: California
;
; COUNTRY: US
;
; ZIP: 94105

```

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patent In Release #1.0, Version #1.25
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/460,510
8  FILING DATE: 02-JUN-1995
9  CLASSIFICATION: 435
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Dow, Karen B.
12 REGISTRATION NUMBER: 29,684
13 REFERENCE/DOCKET NUMBER: 012418-001430
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: (415) 326-2400
16 TELEFAX: (415) 326-2422
17 INFORMATION FOR SEQ ID NO: 5:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 6375 base pairs
20 TYPE: nucleic acid
21 STRANDEDNESS: double
22 TOPOLOGY: linear
23 MOLECULE TYPE: cDNA to mRNA
24 HYPOTHETICAL: NO
25 ANTI-SENSE: NO
26 ORIGINAL SOURCE:
27 ORGANISM: Homo Sapiens
28 STRAIN: lambda gt10
29 FEATURE:
30 NAME/KEY: CDS
31 LOCATION: 129..3395
32 OTHER INFORMATION: /note= "nucleotide number 1 of this
33 OTHER INFORMATION: sequence is identical to the nucleotide number 1
34 OTHER INFORMATION: of the previous 4100 long sequence"
35 US-08-460-510-5

```

Query Match		31.6%;	Score 29.4;	DB 2;	Length 6375;
Best Local Similarity		58.6%;	Pred. No. 8.9;		
Matches	51;	Conservative	0;	Mismatches	36; Indels 0; Gaps 0;
QY	4	agtaacacatatatttgatataaatttactataaaactattttctagtactgttctaactg	63		
Dd	6029	AGTTATACATATACATAAAGATATATCTCAACTCTTATGACGGTTTTCTGAATACTG	6088		
QY	64	tctggaggattttcacagtaataaagaaa	90		
Dd	6089	TTCGACATAGTGACGGAAGCAAATATA	6115		

RESULT 8

RESULT 7

```

US-08-460-490-5
; Sequence 5, Application US/08460490
; Patent No. 5891652
; GENERAL INFORMATION:
; APPLICANT: Wolf, David
; APPLICANT: Tomlinson, James E.
; APPLICANT: Fretto, Larry J.
; APPLICANT: Giese, Neill A.
; APPLICANT: Escobedo, Jaime A.
; APPLICANT: Williams, Lewis T.
; TITLE OF INVENTION: DOMAINS OF EXTRACELLULAR REGION OF HUMAN
; TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR RECEPTOR POLYPEPTIDES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW
; STREET: One Market Plaza, Steuart Street Tower, Suite 2000
; City: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,490
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 012418-001420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6375 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; STRAIN: lambda gt10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 129...3395
; OTHER INFORMATION: /note= "nucleotide number 1 of this
; sequence is identical to the nucleotide number 1
; of the previous 4100 long sequence"
US-08-460-490-5

```

	Query Match	31.6%;	Score 29.4;	DB 2;	Length 6375;
	Best Local Similarity	58.6%;	Pred. No. 8.9;		
	Matches 51;	Conservative 0;	Mismatches 36;	Indels 0;	Gaps 0;
Qy	4 agtaacacataatttgatgaataatttactaaaactattttctcgtagcactgtttaaatcatg	63			
Db	6029 AGTTATACATAATACAAAAAGATATATCTCAACCTCTTATGACGGTTTTTGTAATACTG	6088			
Qy	64 tctgaggattttcacgtaataaaagaaa	90			
Db	6089 TTCGACATAGTCAGCGAAGCAAATATA	6115			

RESULT 8


```

Query Match          31.6%; Score 29.4; DB 5; Length 6375;
Best Local Similarity 58.6%; Pred. No. 8.9;
Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY      4 agtaaacacatatatttgatataaaatttactactaaaactattttctagtactctgtttaatcatg 63
        ||| | ||||| | |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      6029 AGTTATACATATATACATAAAGATATATCTCAACTCTTATGACGGTTTTTGTAATAACTG 6088

QY      64 tctgaggattttcacagtaataaagaaa 90
        || | || | || | || | || | || | || | || | || | || | || | || | || | || |
Db      6089 TTCGACATAGTCAGGAAGCAAAATATA 6115

RESULT 10
US-09-385-982-95
Sequence 95, Application US/09385982

```

```
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 95
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(392)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-95

; Query Match 30.3%; Score 28.2; DB 4; Length 392;
; Best Local Similarity 54.8%; Pred. No. 15;
; Matches 51; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 1 aaagtaacacatttggataaattattactaaactattttctagtaactgttaatc 60
Db 249 aaaaaaagaagtttgtaattttatattacttttaagtgtgatactaaagtattaaac 308
QY 61 atgtctgaggattttacagtaataaagaacga 93
Db 309 atattctgnattctccaaaaaanaanta 341

RESULT 11
US-08-832-883-49/c
; Sequence 49, Application US/08832883
; Patent No. 5807681
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-49

; Query Match 30.3%; Score 28.2; DB 2; Length 4576;
; Best Local Similarity 61.6%; Pred. No. 18;
; Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 6 taacacataattttgataaattattactaaactattttctagtaactgttaatc 65
Db 3477 TTACGTGTAGTTAAACCACTTTTATTATTATCTTAATTTCTAGAAATTAAGAAATTATTC 3418

; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-49
```

```
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-49

; Query Match 30.3%; Score 28.2; DB 1; Length 4576;
; Best Local Similarity 61.6%; Pred. No. 18;
; Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 6 taacacataattttgataaattattactaaactattttctagtaactgttaatc 65
Db 3477 TTACGTGTAGTTAAACCACTTTTATTATTATCTTAATTTCTAGAAATTAAGAAATTATTC 3418

; Query Match 30.3%; Score 28.2; DB 2; Length 4576;
; Best Local Similarity 61.6%; Pred. No. 18;
; Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 6 taacacataattttgataaattattactaaactattttctagtaactgttaatc 65
Db 3477 TTACGTGTAGTTAAACCACTTTTATTATTATCTTAATTTCTAGAAATTAAGAAATTATTC 3418

; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-49
```

```
; Query Match 30.3%; Score 28.2; DB 2; Length 4576;
; Best Local Similarity 61.6%; Pred. No. 18;
; Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY 6 taacacataattttgataaattattactaaactattttctagtaactgttaatc 65
Db 3477 TTACGTGTAGTTAAACCACTTTTATTATTATCTTAATTTCTAGAAATTAAGAAATTATTC 3418
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:18:30 ; Search time 189.71 Seconds

(without alignments)
420.280 Million cell updates/sec

Title: US-09-700-187-2

Perfect score: 93

Sequence: 1 aaagtaacacacatttga.....ttacagtaataaagaacga 93

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	100.0	93	21	AA97384
2	93	100.0	2325	21	AA97385
3	93	100.0	3441	21	AA97382
c 4	33.6	36.1	936	22	AAF58252
c 5	33.6	36.1	936	22	AAF58254
c 6	33.6	36.1	936	22	AAF58257
c 7	33.6	36.1	936	22	AAF58259
c 8	33.6	36.1	936	22	AAF58262
c 9	33.6	36.1	938	22	AAF58255
10	33	35.5	58407	19	AAV21210
11	32.8	35.3	244	22	AAF58238

c	12	32.8	35.3	244	22	AAF58238	Oligonucleotide D1
	13	32.8	35.3	936	22	AAF58252	Oligonucleotide D1
	14	32.8	35.3	936	22	AAF58254	Oligonucleotide D1
	15	32.8	35.3	936	22	AAF58257	Oligonucleotide D1
	16	32.8	35.3	936	22	AAF58259	Oligonucleotide D2
	17	32.8	35.3	936	22	AAF58262	Oligonucleotide D2
	18	32.8	35.3	938	22	AAF58255	Oligonucleotide D1
c	19	32.4	34.8	5535	21	AA70184	Plasmodium falcipa
	20	32.2	34.6	269223	22	AAF28554	Genomic fragment #
c	21	31.8	34.2	3549	21	AA70223	Plasmodium falcipa
c	22	31.6	34.0	441	21	AA70237	human secreted pro
c	23	31	33.3	31	21	AA97411	Pea wild-type pra2
c	24	31	33.3	31	21	AA97412	Pea wild-type pra2
c	25	31	33.3	31	21	AA97423	Pea wild-type pra2
c	26	31	33.3	507	22	AA06274	Domestic mite BtA2
c	27	30.6	32.9	134	16	AA724176	Human gene signatu
c	28	30.2	32.5	1301	21	AA62843	Lipid transfer pro
c	29	30	32.3	1431	21	AA237082	DNA sequence encod
c	30	30	32.3	1671	13	AAQ24134	50 kD subunit of S
c	31	30	32.3	5763	18	AAV74816	Staphylococcus aur
c	32	30	32.3	5924	18	AAV74441	Staphylococcus aur
c	33	29.8	32.0	3001	21	AAH51805	Chromosome 13q31-q
c	34	29.6	31.8	600	18	AA74092	DNA encoding a Sta
c	35	29.4	31.6	877	21	AA54965	Arabidopsis thalia
c	36	29.4	31.6	878	21	AA37095	Arabidopsis thalia
c	37	29.4	31.6	1984	17	AA713030	Cotton fibre-speci
c	38	29.4	31.6	1985	17	AA730250	Cotton fibre clone
c	39	29.4	31.6	1985	18	AA770036	Cotton H6 gene and
c	40	29.4	31.6	2099	21	AA57997	2099 bp Candida al
c	41	29.4	31.6	6075	13	AAQ27451	Type A human plate
c	42	29.4	31.6	129021	21	AAF22296	BAC containing rep
c	43	29.4	31.6	580073	18	AA758840	Mycoplasma genital
c	44	29.4	31.6	910715	20	AA20248	Borrelia burgdorfe
c	45	29.2	31.4	475	21	AAH30428	Human colon cancer

ALIGNMENTS

RESULT 1
AAA97384
ID AAA97384 standard; DNA; 93 BP.
XX
AC AAA97384;
XX
XX
DT 29-JAN-2001 (first entry)
XX
DE Pea pra2 gene light-repressible promoter cis element.
XX
XX GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; cis element; ds.
XX
OS Pisum sativum.
XX
XX WO200055313-A1.
XX
XX 21-SEP-2000.
XX
XX 03-MAR-2000; 2000WO-JP01269.
XX
XX 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR) SUNTORY LTD.
XX
XX Sasaki Y, Nagano Y, Inaba T;
XX
XX WPI; 2000-587526/55.
XX
XX New DNA fragment or promoter for expressing a target gene, specifically
XX under photoinhibitory conditions, and for transforming a plant cell or
XX plant to improve quality and prevent deterioration during storage

PS Claim 2; Page 33; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GTP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. The present sequence
 CC represents the pea pra2 promoter cis element.

XX Sequence 93 BP; 38 A; 10 C; 11 G; 34 T; 0 other;

Query Match 100.0%; Score 93; DB 21; Length 93;
 Best Local Similarity 100.0%; Pred. No. 2.7e-14;
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacacatatatttgataaattattactaaactattttctagctactgttaac 60

Db 1 aaagtaacacacatatatttgataaattattactaaactattttctagctactgttaac 60

QY 61 atgtctgaggattttacagtaataaagaacga 93

Db 61 atgtctgaggattttacagtaataaagaacga 93

RESULT 2

AAA97385

ID AAA97385 standard; DNA; 2325 BP.

XX AC AAA97385;

XX DT 29-JAN-2001 (first entry)

DE Pea pra2 gene light-repressible promoter.

XX GTP-binding protein pra2; pea; light-repressible promoter;

KW photoinhibitory; expression cassette; transgenic plant;

KW deterioration prevention; storage; ds.

XX OS Pisum sativum.

XX PN WO200055313-A1.

XX PD 21-SEP-2000.

XX PF 03-MAR-2000; 2000WO-JP01269.

XX PR 12-MAR-1999; 99JP-0066551.

XX PA (SUNR) SUNTORY LTD.

XX PI Sasaki Y, Nagano Y, Inaba T;

XX PS WPI; 2000-587526/55.

XX The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GTP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. The present sequence
 CC represents the pea pra2 promoter cis element.

XX Claim 3; Page 34-35; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GTP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression

CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. The present sequence
 CC represents the pea pra2 promoter.

SQ Sequence 2325 BP; 791 A; 451 C; 317 G; 766 T; 0 other;

Query Match 100.0%; Score 93; DB 21; Length 2325;

Best Local Similarity 100.0%; Pred. No. 2.8e-14;

Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacacatatatttgataaattattactaaactattttctagctactgttaac 60

Db 1396 aaagtaacacacatatatttgataaattattactaaactattttctagctactgttaac 1455

QY 61 atgtctgaggattttacagtaataaagaacga 93

Db 1456 atgtctgaggattttacagtaataaagaacga 1488

RESULT 3

AAA97382

ID AAA97382 standard; cDNA; 3441 BP.

XX AC AAA97382;

XX DT 29-JAN-2001 (first entry)

DE Pea light-repressible GTP-binding protein pra2 cDNA.

XX GTP-binding protein pra2; pea; light-repressible promoter;

KW photoinhibitory; expression cassette; transgenic plant;

KW deterioration prevention; storage; ss.

XX OS Pisum sativum.

XX PN WO200055313-A1.

XX PD 21-SEP-2000.

XX PF 03-MAR-2000; 2000WO-JP01269.

XX PR 12-MAR-1999; 99JP-0066551.

XX PA (SUNR) SUNTORY LTD.

XX PI Sasaki Y, Nagano Y, Inaba T;

XX PS WPI; 2000-587526/55.

XX P-PSDB; AAB23171.

XX New DNA fragment or promoter for expressing a target gene, specifically
 CC under photoinhibitory conditions, and for transforming a plant cell or
 CC plant to improve quality and prevent deterioration during storage -
 CC Example 1; Fig 1; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or

CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GTP-binding protein pra2. The invention also relates to an
 CC expression cassette containing the pra2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. The present sequence
 CC represents the pea pra2 gene.

SQ Sequence 3441 BP; 1143 A; 606 C; 540 G; 1152 T; 0 other;

```

Query Match      100.0%; Score 93; DB 21; Length 3441;
Best Local Similarity 100.0%; Pred. NO. 2.8e-14;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aaagtaacacataattttgataaaattttattactaaaactattttctagtagtctgttaac 60
    |||||
Db 1396 aaagtaacacataattttgataaaattttattactaaaactattttctagtagtctgttaac 1455
    |||||

QY 61 atgtctgaggattttacagtaataaagaacga 93
    |||||
Db 1456 atgtctgaggattttacagtaataaagaacga 1488
    |||||

RESULT 4
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PS 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match      36.1%; Score 33.6; DB 22; Length 936;
Best Local Similarity 3.3%; Pred. NO. 6.1;
Matches 3; Conservative 69; Mismatches 18; Indels 0; Gaps 0;

QY 1 aaagtaacacataattttgataaaattttattactaaaactattttctagtagtctgttaac 60
    :::::
Db 461 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 402

QY 61 atgtctgaggattttacagtaataaagaacga 90
    :::::
Db 401 WWWWWW TAAGC WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 372

RESULT 5
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX

```

```

AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PS 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match      36.1%; Score 33.6; DB 22; Length 936;
Best Local Similarity 3.3%; Pred. NO. 6.1;
Matches 3; Conservative 69; Mismatches 18; Indels 0; Gaps 0;

QY 1 aaagtaacacataattttgataaaattttattactaaaactattttctagtagtctgttaac 60
    :::::
Db 461 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 402

QY 61 atgtctgaggattttacagtaataaagaacga 90
    :::::
Db 401 WWWWWW TAAGC WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 372

RESULT 6
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX

```



```

XX OS Synthetic.
XX SS WO200107665-A2.
XX PN
XX PD 01-FEB-2001.
XX PF
XX PR 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -
XX PS Example 4; Page 120; 159pp; English.
XX CC The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.
XX SQ Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;

Query Match      35.3%; Score 32.8; DB 22; Length 244;
Best Local Similarity 1.1%; Pred. No. 9.3;
Matches 1; Conservative 71; Mismatches 18; Indels 0; Gaps

Qy   1 aaagtaacacatatcttgataaaatttactaaaacctatttttagtactgtgaatc 60
    :::: :: : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db   191 WWWWWWWWXXXXXXXXXXXXXXXWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 132
    :::: :: : ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy   6l atgctgaggagtattcacagtaagaaga 90
    : : : : : ::::: : :::::|:::
Db   131 WWWWWWWWXXXXXXXXXXXXXXXWWWWWWWWWWWWWWWWWWWWWWWWWWWW 102
    : : : : : ::::: : :::::|:::

RESULT 13
AAF58252
ID   AAF58252 standard; DNA; 936 BP.
XX AC
XX AC AAF58252;
XX DT
XX DE 24-APR-2001 (first entry)
XX DE Oligonucleotide D1835.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
KW KW gene expression; ss.
XX OS Synthetic.
XX OS
PN   WO200107665-A2.
XX PD
XX PD 01-FEB-2001.
XX PF
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PA
XX XX

```


Qy 61 atgtctgaggattttacagtaataaagaaa 90
:: : : : : : : : : : : : : : : : : :
Db 82 wwwwwwwwwwwwwwwwwwwwwwwwwww 111

Search completed: April 3, 2002, 05:18:34
Job time: 2904 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:18:34 ; Search time 189.71 Seconds
(without alignments)
10506.991 Million cell updates/sec

Title: US-09-700-187-3

Perfect score: 2325

Sequence: 1 aagctttaaggcaaggaa.....ttttgatcttgacagaaga 2325

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 42862619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2325	100.0	2325	21 AAA97385	Pea pra2 gene ligh
2	2325	100.0	3441	21 AAA97382	Pea light-repressi
C 3	198.2	8.5	936	22 AAF58252	Oligonucleotide D1
C 4	198.2	8.5	936	22 AAF58254	Oligonucleotide D1
C 5	198.2	8.5	936	22 AAF58257	Oligonucleotide D1
C 6	198.2	8.5	936	22 AAF58259	Oligonucleotide D2
C 7	198.2	8.5	936	22 AAF58262	Oligonucleotide D2
C 8	198.2	8.5	936	22 AAF58255	Oligonucleotide D1
9	197.6	8.5	936	22 AAF58252	Oligonucleotide D1
10	197.6	8.5	936	22 AAF58254	Oligonucleotide D1
11	197.6	8.5	936	22 AAF58257	Oligonucleotide D1

12	197.6	8.5	936	22 AAF58259	Oligonucleotide D2
13	197.6	8.5	936	22 AAF58262	Oligonucleotide D2
14	197.6	8.5	938	22 AAF58255	Oligonucleotide D1
15	93	4.0	93	21 AAA97384	Pea pra2 gene ligh
16	75.4	3.2	244	22 AAF58238	Oligonucleotide D1
C 17	71.8	3.1	244	22 AAF58238	Oligonucleotide D1
18	53.2	2.3	5409	21 AAF70151	Plasmodium falcipa
19	52.4	2.3	513445	22 AA161373	Soybean 318013 reg
20	51.4	2.2	1664976	19 AAV21209	Methanococcus jann
21	50.4	2.2	19124	18 AAT72882	Plasmodium var-7 g
22	50.4	2.2	19124	21 AAZ98287	Plasmodium var-7 p
23	50.2	2.2	1627	14 AAQ47761	Antisense RNA comp
C 24	50.2	2.2	6152	18 AAT78867	P. falciparum live
25	49.4	2.1	1711	19 AAV33136	Plasmodium berghei
C 26	49.2	2.1	1230	10 AAN90223	Malaria-specific D
27	48.4	2.1	9789	17 AAT41852	cDNA encoding Plas
28	48.2	2.1	3975	9 AAN81157	Malaria-specific g
29	48.2	2.1	3975	13 AAQ22999	SERP gene. Plasm
30	48.2	2.1	6124	11 AAQ03568	Sequence encoding
31	47.2	2.0	1132	21 AAF12929	Aspergillus oryzae
C 32	46.8	2.0	2030	19 AAV29218	Nucleotide sequenc
C 33	46.4	2.0	3101	11 AAQ02047	Sequence encoding
C 34	46.4	2.0	10409	19 AAV42558	Mouse dectin-2 gen
35	46.2	2.0	441	21 AAC69763	Human breast tumo
36	46.2	2.0	53585	20 AAQ20251	Borrelia burgdorfe
C 37	46	2.0	488	21 AAC94232	Cat flea head and
38	46	2.0	161425	22 AAH02340	Human AKAP10 gene
39	46	2.0	162025	22 AAH02339	Human AKAP10 gene
40	45.8	2.0	376	22 AAH93354	Human chromosome 1
41	45.8	2.0	1431	21 AAZ37082	DNA sequence encod
42	45.8	2.0	1671	13 AAQ24134	50 kD subunit of S
43	45.8	2.0	3285	22 AAH54777	S. epidermidis gen
C 44	45.8	2.0	3314	22 AAH54563	S. epidermidis gen
C 45	45.6	2.0	1078	21 AAQ02581	Human colon cancer

ALIGNMENTS

RESULT 1
AAA97385
ID AAA97385 standard; DNA; 2325 BP.
XX
AC AAA97385;
XX
DT 29-JAN-2001 (first entry)
XX
DE Pea pra2 gene light-repressible promoter.
XX
KW GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; ds.
XX
OS Pisum sativum.
XX
PN WO200055313-A1.
XX
PD 21-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-JP01269.
XX
PR 12-MAR-1999; 99JP-0066551.
XX
XX (SUNR) SUNTORY LTD.
XX
PI Sasaki Y, Nagano Y, Inaba T;
XX
DR WPI; 2000-587526/55.
XX
PT New DNA fragment or promoter for expressing a target gene, specifically
PT under photoinhibitory conditions, and for transforming a plant cell or
PT plant to improve quality and prevent deterioration during storage -

PS Claim 3; Page 34-35; 49pp; Japanese.

XX The invention relates to a light-repressible promoter (AAA97385), or
CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
CC the pea GTP-binding protein pra2. The invention also relates to an
CC expression cassette containing the pra2 promoter or its active
CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. The present sequence
XX represents the pea pra2 promoter.

SQ Sequence 2325 BP; 791 A; 451 C; 317 G; 766 T; 0 other;

Query Match 100.0%; Score 2325; DB 21; Length 2325;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	aagctttaagcgcaagggaagacacaattccaaaataataaaaactcctaaagaatga	60
DB	1	aagctttaagcgcaagggaagacacaattccaaaataataaaaactcctaaagaatga	60
QY	61	ttttattcttattcctaataaacttttctattccaaaaacacatcaaaagtattgtga	120
DB	61	ttttattcttattcctaataaacttttctattccaaaaacacatcaaaagtattgtga	120
QY	121	ttcatacttttaattctgataataataattgtattatattcaataattttcatacaattgtg	180
DB	121	ttcatacttttaattctgataataataattgtattatattcaataattttcatacaattgtg	180
QY	181	ttatatgaataatttttgagttaaaaggactaagaataacctccgcaacatcaaaagtca	240
DB	181	ttatatgaataatttttgagttaaaaggactaagaataacctccgcaacatcaaaagtca	240
QY	241	gaacaccttgtaacctcttcagttgaaacgagagaagtggaacacacagaaactaaag	300
DB	241	gaacaccttgtaacctcttcagttgaaacgagagaagtggaacacacagaaactaaag	300
QY	301	ttccccacttaacctcttggttggtgagagactcccttcaaatatttactctaagga	360
DB	301	ttccccacttaacctcttggttggtgagagactcccttcaaatatttactctaagga	360
QY	361	aatacatagacactctagatgggttgcatagctacatatattttaagtaataataccc	420
DB	361	aatacatagacactctagatgggttgcatagctacatatattttaagtaataataccc	420
QY	421	acttcaagttttttgtttttgttggtagcagtagatgataagatggatcatttctcaa	480
DB	421	acttcaagttttttgtttttgttggtagcagtagatgataagatggatcatttctcaa	480
QY	481	ggcccttatgcaagacataaagatccatatactccaccaagattgtttcatactcaacca	540
DB	481	ggcccttatgcaagacataaagatccatatactccaccaagattgtttcatactcaacca	540
QY	541	agttaatgaatttcaattcttcgaacaaatttttctcaccagaaggaagtattatgcac	600
DB	541	agttaatgaatttcaattcttcgaacaaatttttctcaccagaaggaagtattatgcac	600
QY	601	attttctaatgtatttttatataagaattgatacatgttttctgttatacaagaattagaatt	660
DB	601	attttctaatgtatttttatataagaattgatacatgttttctgttatacaagaattagaatt	660
QY	661	tggatttctcaatccaaactcctacacttggtagaataatttcagacctcaacctcagtaaat	720
DB	661	tggatttctcaatccaaactcctacacttggtagaataatttcagacctcaacctcagtaaat	720
QY	721	caggttctctcttcaaaactcatcaacttgggttagtgagagaatttatggacgtcaacctagc	780
DB	721	caggttctctcttcaaaactcatcaacttgggttagtgagagaatttatggacgtcaacctagc	780

QY	781	aatatgaatccctctccatccagatccctacacttatctgtagtgagaaattttgtctcctcgacct	840
DB	781	aatatgaatccctctccatccagatccctacacttatctgtagtgagaaattttgtctcctcgacct	840
QY	841	caacaagatagatttgatgggtcatcacgagggggaagcattcacattgggttcaaaagattc	900
DB	841	caacaagatagatttgatgggtcatcacgagggggaagcattcacattgggttcaaaagattc	900
QY	901	accacaacaagtggagagacatcacatatacaaccaaaacttaagtgtagtgatg	960
DB	901	accacaacaagtggagagacatcacatatacaaccaaaacttaagtgtagtgatg	960
QY	961	agttctcttactataaaagtgtcaacctccacttttcttaagcaattgtgacttagaac	1020
DB	961	agttctcttactataaaagtgtcaacctccacttttcttaagcaattgtgacttagaac	1020
QY	1021	tcacacttatttctcaacaataaactcacacttgttttatacaacaattcccccaagtg	1080
DB	1021	tcacacttatttctcaacaataaactcacacttgttttatacaacaattcccccaagtg	1080
QY	1081	agttcattcgtctatgtccctcccaagtggaaatctcttctcccgcatgcttataccggtg	1140
DB	1081	agttcattcgtctatgtccctcccaagtggaaatctcttctcccgcatgcttataccggtg	1140
QY	1141	ttgacatacatcttactctgctcgtggcacttcaatgggacaogctcctgaccaccatg	1200
DB	1141	ttgacatacatcttactctgctcgtggcacttcaatgggacaogctcctgaccaccatg	1200
QY	1201	tcagaagaacttttgacacaagaggtcggtcccttactcgaaacagactctgataccatt	1260
DB	1201	tcagaagaacttttgacacaagaggtcggtcccttactcgaaacagactctgataccatt	1260
QY	1261	aatagatcaactttgaaatgggatatcattcactatactatacaaacatttactgtaaaagataaaa	1320
DB	1261	aatagatcaactttgaaatgggatatcattcactatactatacaaacatttactgtaaaagataaaa	1320
QY	1321	aatcccccaaaaatggagagagacactacatctctcttattataataataaaatgttaa	1380
DB	1321	aatcccccaaaaatggagagagacactacatctctcttattataataataaaatgttaa	1380
QY	1381	agaaaaatagatataaaagtaacacatatatttgataaaatttactataaacatttttc	1440
DB	1381	agaaaaatagatataaaagtaacacatatatttgataaaatttactataaacatttttc	1440
QY	1441	tagtacttgttaatactgtctgaggattttcacagtaataaagaacagaggttagcccaaac	1500
DB	1441	tagtacttgttaatactgtctgaggattttcacagtaataaagaacagaggttagcccaaac	1500
QY	1501	aaaagtgataattgtgaggggtgtgactcttctgtcgggtgcaaaaaaatgaaaccccaactt	1560
DB	1501	aaaagtgataattgtgaggggtgtgactcttctgtcgggtgcaaaaaaatgaaaccccaactt	1560
QY	1561	gtgatatgtgtcgactgtccgtcgtacattgaaatttaattgttcttttataacg	1620
DB	1561	gtgatatgtgtcgactgtccgtcgtacattgaaatttaattgttcttttataacg	1620
QY	1621	tttgtctatgcgttattccccatattgtctacatagaattgggacaatgaatttataatat	1680
DB	1621	tttgtctatgcgttattccccatattgtctacatagaattgggacaatgaatttataatat	1680
QY	1681	ctgtcatgtgtgggtggatccaatttaattgtatcgtaaatggtaggacatactatgct	1740
DB	1681	ctgtcatgtgtgggtggatccaatttaattgtatcgtaaatggtaggacatactatgct	1740
QY	1741	acacaattatcatcatcactgtgcaactcacgtgcaattgttcttctccatgaattc	1800
DB	1741	acacaattatcatcatcactgtgcaactcacgtgcaattgttcttctccatgaattc	1800
QY	1801	acattgtctaaagaaaaattaccaccttaaaattgtttatcccttgcacacattttcacatcaa	1860
DB	1801	acattgtctaaagaaaaattaccaccttaaaattgtttatcccttgcacacattttcacatcaa	1860
QY	1861	tttattaaaacatttttaccatttggaaaaacacatactattcaatcaattttttgcaatt	1920

[illegible]

RESULT 2
AAA97382

ID AAA97382 standard; cDNA; 3441 BP.

XX
AC AAA97382.

XX
DT 29-JAN-2001 (first entry)

XX DE pea light-repressible GTP-binding protein pra2 cDNA.

AA GTP-binding protein pra2; pea; light-repressible promoter;
KW photoinhibitory; expression cassette; transgenic plant;
KW deterioration prevention; storage; ss.
KW

XX
OS Pisum sativum.

XX PN WO200055313-A1.

XX
PD 21-SEP-2000.XX
PF
03-MAR-2000;XX
PR 12-MAR-1999; 99JP-0066551.

XX PA (SUNR) SUNTORY LTD.

XX
PI Sasaki Y, Nagano Y,

XX DR WPI; 2000-587526/55.

DR P-PSDB; AAB231/L.
XX

PT new DNA fragment or promoter for expressing a target gene, specifically under photoinhibitory conditions, and for transforming a plant cell or plant to improve quality and prevent deterioration during storage -

AA
PS
Example 1; Fig 1; 49pp; Japanese.

The invention relates to a light-repressible promoter (AAA97385), or active fragments thereof (AAA97383, AAA97384), from the gene encoding the pea GTP-binding protein pr2. The invention also relates to an expression cassette containing the pr2 promoter or its active

CC fragments for the expression of a gene under photoinhibitory or dark
CC conditions in a plant, and to transgenic plants, their descendants
CC and plant tissues comprising the expression cassette. The expression
CC cassette of the invention can be used to generate transgenic plants in
CC which deterioration during storage in the dark is prevented. This is
CC particularly useful for agricultural products. The present sequence
CC represents the pea praz2 gene.

Sequence 3441 BP; 1143 A; 606 C; 540 G; 1152 T; 0 other; XX

Query Match	100.0%;	Score 2325;	DB 21;	Length 3441;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2325;	Conservative	0;	Mismatches	0; Indels 0;

Qy	1	aagctttaaggcaaggaaagacaacaactccaataataaaaaataaaaaactcctaagaatga	60
Db	1	aagctttaaggcaaggaaagacaacaactccaataataaaaaataaaaaactcctaagaatga	60
Qy	61	ttttattcttatcttcataaaataacttttcttatctccaataaaacacatcaaatgttatgtga	120
Db	61	ttttattcttatcttcataaaataacttttcttatctccaataaaacacatcaaatgttatgtga	120
Qy	121	ttcatatctttaattatctgataaataataattgtatatccaattttcatacaaatgtgt	180
Db	121	ttcatatctttaattatctgataaataataattgtatatccaattttcatacaaatgtgt	180
Qy	181	tttatgaaaatattttgtaggtaaaaggactaaagaataaacttcgcacatcaaatgca	240
Db	181	tttatgaaaatattttgtaggtaaaaggactaaagaataaacttcgcacatcaaatgca	240
Qy	241	gaaacctctgtgaactcttcagtgaacgagaagaagtggaacaacagaaaaactaaag	300
Db	241	gaaacctctgtgaactcttcagtgaacgagaagaagtggaacaacagaaaaactaaag	300
Qy	301	ttcccccacttaactcttggtttgggtgaggaacttcctttacaaatttatactctaagga	360
Db	301	ttcccccacttaactcttggtttgggtgaggaacttcctttacaaatttatactctaagga	360
Qy	361	aatacatatgagaactctagatgggttgcattagctctaataatttttaagtaataataccc	420
Db	361	aatacatatgagaactctagatgggttgcattagctctaataatttttaagtaataataccc	420
Qy	421	acttcaagttttttgtttttgttggctgagtagatgaataagatgcatctttctcaa	480
Db	421	acttcaagttttttgtttttgttggctgagtagatgaataagatgcatctttctcaa	480
Qy	481	ggcccttatgcaagacataaagatccatactccaccaagaattgctttacatcttaacca	540
Db	481	ggcccttatgcaagacataaagatccatactccaccaagaattgctttacatcttaacca	540
Qy	541	agtaatagtaatttaattcttcgaacaataatttctccaccaagggaagtttatatgcac	600
Db	541	agtaatagtaatttaattcttcgaacaataatttctccaccaagggaagtttatatgcac	600
Qy	601	attttctaatgtattttatatagaattgatactgtttctgtttatatacaagattagaatt	660
Db	601	attttctaatgtattttatatagaattgatactgtttctgtttatatacaagattagaatt	660
Qy	661	tggattttctcatccaaactcctacacttggtagagaaatttcagctccaactcagtaaat	720
Db	661	tggattttctcatccaaactcctacacttggtagagaaatttcagctccaactcagtaaat	720
Qy	721	caggttctctcttcaaaactcatacacttgggttgagtgaagaattatgagctccaactgac	780
Db	721	caggttctctcttcaaaactcatacacttgggttgagtgaagaattatgagctccaactgac	780
Qy	781	aataatgaatccctctccaagatcctacacttatctgagtgagaaatttgggtcctcgacct	840
Db	781	aataatgaatccctctccaagatcctacacttatctgagtgagaaatttgggtcctcgacct	840
Qy	841	caacaagatagattttagtggttcataccagggggaagcattcaacattgggttcaagaattc	900

Db 841 caacaagatagattggtggtcaccagggggaagcattcacattggtggtcaagattc 900
Qy 901 acccaacaagtgagagagacacacatcaacacaaaccccttaagtgataggtgatg 960
Db 901 acccaacaagtgagagagacacacatcaacacaaaccccttaagtgataggtgatg 960
Qy 961 agttctcttacttaaaagtgtccaacctccacttttcttaagcaatgtgtgacttagaac 1020
Db 961 agttctcttacttaaaagtgtccaacctccacttttcttaagcaatgtgtgacttagaac 1020
Qy 1021 tcacaactattctcaacaactcaactcaactgttttatcaacaatctccccacaagtgtg 1080
Db 1021 tcacaactattctcaacaactcaactgttttatcaacaatctccccacaagtgtg 1080
Qy 1081 agttcaatcgctatgtccccctcaagtgaactctcttcatcgcgatgcttataccgttg 1140
Db 1081 agttcaatcgctatgtccccctcaagtgaactctcttcatcgcgatgcttataccgttg 1140
Qy 1141 ttgacatacatotttactcgttcgacttcaatgggacacgctgctgaccaccatg 1200
Db 1141 ttgacatacatotttactcgttcgacttcaatgggacacgctgctgaccaccatg 1200
Qy 1201 tcaagaagacttttgacacaagagtcggtcccttactcgaaccagactctgataccatt 1260
Db 1201 tcaagaagacttttgacacaagagtcggtcccttactcgaaccagactctgataccatt 1260
Qy 1261 aatagatcactttgaatgatcattcattactatcatcaaacatttacgtaaagataaaa 1320
Db 1261 aatagatcactttgaatgatcattcattactatcatcaaacatttacgtaaagataaaa 1320
Qy 1321 aattcaccccaaaaatgagagacacacacatctctcttatttatataataaaatgtaa 1380
Db 1321 aattcaccccaaaaatgagagacacacacatctctcttatttatataataaaatgtaa 1380
Qy 1381 agaaaaatagtaataaaagtaaacacatttttgataaaatttattactaaaactatttc 1440
Db 1381 agaaaaatagtaataaaagtaaacacatttttgataaaatttattactaaaactatttc 1440
Qy 1441 tagtactgttaataatgctcgtgaggtttacagtaataaagaacgaggtagcccaaac 1500
Db 1441 tagtactgttaataatgctcgtgaggtttacagtaataaagaacgaggtagcccaaac 1500
Qy 1501 aaaagtataatgtgaggggtgtagcttctgtcgtgcaaaaaatgaaacccccaaactt 1560
Db 1501 aaaagtataatgtgaggggtgtagcttctgtcgtgcaaaaaatgaaacccccaaactt 1560
Qy 1561 gtgatatgtgtcgaactcgtcgtcgtcattgaaatttaagaaatgttcttttataacg 1620
Db 1561 gtgatatgtgtcgaactcgtcgtcgtcattgaaatttaagaaatgttcttttataacg 1620
Qy 1621 ttgtctatgccgtattaccatattgctcactagatgggacaatgaatttaatatatat 1680
Db 1621 ttgtctatgccgtattaccatattgctcactagatgggacaatgaatttaatatatat 1680
Qy 1681 ctgtcatgtgtgggtggtatcaatttaattgtatcgttaaatggttaggacatactcatgct 1740
Db 1681 ctgtcatgtgtgggtggtatcaatttaattgtatcgttaaatggttaggacatactcatgct 1740
Qy 1741 acacaattatatacactcgttcaatcactcgtcgaatgtgtttctcttcccatgaattc 1800
Db 1741 acacaattatatacactcgttcaatcactcgtcgaatgtgtttctcttcccatgaattc 1800
Qy 1801 acattgctaaaagaaattaccaccttaaaatgtttatcccttgacacatttccacatcaa 1860
Db 1801 acattgctaaaagaaattaccaccttaaaatgtttatcccttgacacatttccacatcaa 1860
Qy 1861 ttatttaaacaattttaccattggaacacacacacatatttcaatcaattattttgtcatt 1920
Db 1861 ttatttaaacaattttaccattggaacacacacacatatttcaatcaatattttgtcatt 1920
Qy 1921 ttcaaaaactaaacaaacaaacttagaatttttgaattatttagcacaaattttcaaaaa 1980
Db 1921 ttcaaaaactaaacaaacaaacttagaatttttgaattatttagcacaaattttcaaaaa 1980

Qy 1981 tatcttagtcttcaaccactcaataattcacaaattttccaaatcccttgcaaaacatcaca 2040
Db 1981 tatcttagtcttcaaccactcaataattcacaaattttccaaatcccttgcaaaacatcaca 2040
Qy 2041 acctctagaaaactttgatttaataatctaaataaaagcaataatgtatctaaacaatat 2100
Db 2041 acctctagaaaactttgatttaataatctaaataaaagcaataatgtatctaaacaatat 2100
Qy 2101 caccatatagtttatgataataatgatgacgaataacacttaacttaattgttaagaacttaa 2160
Db 2101 caccatatagtttatgataataatgatgacgaataacacttaacttaattgttaagaacttaa 2160
Qy 2161 agcggagacaactcttataaacacgcgttaattcaacaacgcgttgttcgaggttcattgttt 2220
Db 2161 agcggagacaactcttataaacacgcgttaattcaacaacgcgttgttcgaggttcattgttt 2220
Qy 2221 ctccaactctttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2280
Db 2221 ctccaactctttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2280
Qy 2281 atatactatctctcttgaaactctttttgtatcttggacaagaaa 2325
Db 2281 atatactatctctcttgaaactctttttgtatcttggacaagaaa 2325

RESULT 3
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.

XX AC AAF58252;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1835.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX OS gene expression; ss.
XX OS Synthetic.
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX Umek RM;
XX WPI; 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface
XX Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 8.5%; Score 198.2; DB 22; Length 936;

[illegible]

Qy	1529	tttgcgtgcacaaaaatgaaaccccaacttgcattgtgcagctgcgcgtcgt	1589
Db	183	tttgcgtgcacaaaaatgaaaccccaacttgcattgtgcagctgcgcgtcgt	242
Qy	1589	acattgaaattaatgaattctttttatataacgtttgtctatgcggtattacccatatggt	1648
Db	243	acattgaaattaatgaattctttttatataacgtttgtctatgcggtattacccatatggt	302
Qy	1649	cactagaatggacaatgaatttaatatatactctgcctgctgtggtggattccaatttaa	1708
Db	303	cactagaatggacaatgaatttaatatatactctgcctgctgtggtggattccaatttaa	362
Qy	1709	ttgtatgctaaatggtgagacatactcatgctcacaaattatatcatcactgcgttcacatca	1768
Db	363	ttgtatgctaaatggtgagacatactcatgctcacaaattatatcatcactgcgttcacatca	422
Qy	1769	ctggtcaatgtgtttctctctccatgaattcacattgctcaaaagaaattaccaccccttaa	1828
Db	423	ctggtcaatgtgtttctctctccatgaattcacattgctcaaaagaaattaccaccccttaa	482
Qy	1829	aatggtttcccttgcacacatttcacatcaatttattaaaaacattttaccatttggaataa	1888
Db	483	aatggtttcccttgcacacatttcacatcaatttattaaaaacattttaccatttggaataa	542
Qy	1889	cacatacatatccaatcaattatttttgcattttcaaaaaactaaacacaaacttaga	1948
Db	543	cacatacatatccaatcaattatttttgcattttcaaaaaactaaacacaaacttaga	602
Qy	1949	atattttgtaattatagcacaattttcaaaaataatcctagctcttcaaccactccaataatt	2008
Db	603	atattttgtaattatagcacaattttcaaaaataatcctagctcttcaaccactccaataatt	662
Qy	2009	cacaaattccaaatcccttgcacaaacatcacacacctctagaaaactttgattaaatactta	2068
Db	663	cacaaattccaaatcccttgcacaaacatcacacacctctagaaaactttgattaaatactta	722
Qy	2069	ataaagcacaataatgatcatctaaacaatacacaataatgtttatgataataatgat	2128
Db	723	ataaagcacaataatgatcatctaaacaatacacaataatgtttatgataataatgat	782
RESULT 10			
AAAF58254			
ID AAF58254 standard; DNA; 936 BP.			
AAAF58254;			
24-APR-2001 (first entry)			
Oligonucleotide D1875.			
Electron-transfer group; ETM; mismatch; genotyping;			
gene expression; ss.			
Synthetic.			
WO200107665-A2.			
01-FEB-2001.			
26-JUL-2000; 2000WO-US20476.			
26-JUL-1999; 99US-0145695.			
17-MAR-2000; 2000US-0190259.			
(CLIN-) CLINICAL MICRO SENSORS INC.			
Umek RM;			
WPI; 2001-159728/16.			
Nucleic acids containing electron-transfer group, useful as labels in			
hybridization assays, e.g. for genotyping, allowing repeat analyses on			

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on

[illegible]

RESULT 11
AAF58257

[illegible]

Search completed: April 3, 2002, 05:19:04
Job time: 2934 sec

XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 05:14:55 ; Search time 81.95 Seconds
(without alignments)
6425.398 Million cell updates/sec

Title: US-09-700-187-3
Perfect score: 2325
Sequence: 1 aagctttaaggaaggaa.....ttttgatcttgacaagaaa 2325

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.4	2.2	19124	2	US-08-487-826B-13
c 2	50.2	2.2	6152	2	US-08-973-462-1
c 3	46.8	2.0	2030	2	US-08-705-937-7
c 4	46.4	2.0	10409	3	US-08-772-440-33
5	45.6	2.0	2861	1	US-08-299-953-1
6	45.6	2.0	2861	1	US-08-459-415-1
7	45.6	2.0	2861	5	PCT-US95-11231-1
8	45.6	2.0	3881	1	US-08-299-953-2
9	45.6	2.0	3881	1	US-08-459-415-2
10	45.6	2.0	3881	5	PCT-US95-11231-2
c 11	44.6	1.9	731	1	US-08-451-405A-2
c 12	44.2	1.9	1183	2	US-08-731-722-8
c 13	44	1.9	854	4	US-08-998-416-534
c 14	44	1.9	19124	2	US-08-487-826B-13
c 15	43.8	1.9	5852	1	US-07-867-106-2
c 16	43.6	1.9	1422	1	US-08-319-704-5
c 17	43.2	1.9	1850	3	US-08-617-860B-32
c 18	43.2	1.9	4098	2	US-08-605-106-4
c 19	43	1.8	665	2	US-08-883-795A-36
c 20	43	1.8	2781	3	US-08-749-522-4
c 21	43	1.8	7218	1	US-08-232-463-14
c 22	42.8	1.8	615	4	US-08-998-416-186
c 23	42.8	1.8	827	4	US-08-998-416-535
c 24	42.6	1.8	662	4	US-08-998-416-185
c 25	42.6	1.8	663	4	US-08-998-416-191
c 26	42.6	1.8	665	4	US-08-998-416-937
c 27	42.6	1.8	701	4	US-08-998-416-701

c 28	42.6	1.8	711	4	US-08-998-416-786	Sequence 786, App
c 29	42.6	1.8	724	4	US-08-998-416-683	Sequence 683, App
c 30	42.6	1.8	732	4	US-08-998-416-1036	Sequence 1036, App
c 31	42.6	1.8	767	4	US-08-998-416-472	Sequence 472, App
c 32	42.6	1.8	828	4	US-08-998-416-538	Sequence 538, App
c 33	42.6	1.8	834	4	US-08-998-416-305	Sequence 305, App
c 34	42.6	1.8	860	4	US-08-998-416-287	Sequence 287, App
c 35	42.4	1.8	636	4	US-08-998-416-1137	Sequence 1137, App
c 36	42.4	1.8	658	4	US-08-998-416-595	Sequence 595, App
c 37	42.2	1.8	615	4	US-08-998-416-186	Sequence 186, App
c 38	42.2	1.8	834	4	US-08-998-416-305	Sequence 305, App
c 39	42	1.8	1431	4	US-09-316-083-2	Sequence 2, Appli
c 40	41.8	1.8	5394	3	US-08-688-376-1	Sequence 1, Appli
c 41	41.6	1.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
c 42	41.6	1.8	8920	4	US-09-150-741-1	Sequence 1, Appli
c 43	41.4	1.8	2287	4	US-08-845-258-8	Sequence 8, Appli
c 44	41.4	1.8	2287	4	US-08-990-571-8	Sequence 8, Appli
c 45	41.4	1.8	2287	4	US-08-723-142A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-487-826B-13
; Sequence 13, Application us/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match

2.2%; Score 50.4; DB 2; Length 19124;

```
Best Local Similarity 48.0%; Pred. No. 0.038;
Matches 144; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 1831 ttttttcccttgccacatttcacatcaattttattataaaacattttacatttggaanaa 1890
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18094 tcttactcgaatttcgaatttcgacgagcgaacaaataatataatctcattataaaatt 18153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1891 catacatatcaatcaattatttttgcatcttcaaaacactcaaaccaacaaacttagaat 1950
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18154 atttataatcaatattatattatttcttcttatttataaataatttataatataatataat 18213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1951 attttgtaattagcacaaatttttcaaaataatccttagctcttcaacccactcaataattca 2010
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18214 atttcttggtatttttataaataataactaatttcttatttttttatttataactttattcccttt 18273
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2011 caatttccaaatcccttgcaaaacatcacaaacctctagaaacttgattgaataatccta 2070
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18274 ttaatttcttaattcttttatcaaaacaaacaaacataaagtaatttctacataatcaacaaaa 18333
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2071 aaaaacaaataatgatattcaaaacaaatcacccatatattgtatgatataatgatgc 2130
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18334 aaaaaaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 18393
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-08-973-462-1/c
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

Query Match 2.2%; Score 50.2; DB 4; Length 6152;
Best Local Similarity 45.8%; Pred. No. 0.031;
Matches 210; Conservative 0; Mismatches 248; Indels 1; Gaps 1;

QY 1809 aaagaaaattaccaccttaaaatgtttatcccttgccacacatttcacatcaattttatttaa 1868
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6052 aaaaaaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 1993
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1869 aacattttaccattggaacacacacacacacacacacacacacacacacacacacacacacac 1928
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5992 ttctaatatataatataatataatataatataatataatataatataatataatataatataat 5933
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1929 ctaaacacaaacaaacttagaattttgttaatttagcacaaattttcacaataatccttag 1988
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5932 acaaaataattataattatataatataatataatataatataatataatataatataatataat 5873
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1989 tttcaaacactcaataattcacaatttcacaataatcccttgcaaacacacacacacacacac 2048
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5872 atatttaattta-aaaaacataaaactctgttaaaacacattttataatttaaaataatttg 5814
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2049 aaacttttgatttaattctataaagaacataatgatgatataacacataatcaccattat 2108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5813 aaaaataaattatattatgcgaacaaataatataatataatataatataatataatataatata 5754
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2109 atgttatgataataatgatgcagcaacacacttaatttggttaagcatttaagcgagac 2168
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 5753 AATTTTGTAAAAAGATACATATATATATATATATATATATATATATATATATATATATTTT 5694
QY 2169 aactctattaaacacggtaattcaacaaccggtgtgtgtgcagagttcattgtttcttccaac 2228
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5693 ATATATATTTTACATATATATATATATATATATATATATATATATATATATATATATTTT 5634
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2229 tctttcccttcccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2267
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5633 TCCATCATTTTCACCTCTCTGCAATTTTATTTTGAAT 5595
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

; Sequence 7, Application US/08705937
; Patent No. 5981841
; GENERAL INFORMATION:
; APPLICANT: Santino, Colleen G.
; APPLICANT: Conner, Timothy W.
; TITLE OF INVENTION: EARLY SEED 5' REGULATORY SEQUENCE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/705,937
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicolas G. Barzoukas.
; REGISTRATION NUMBER: 38,823
; REFERENCE/DOCKET NUMBER: MOBT:018 (38-2(10694))A
; NAME: Janelle D. Waack.
; REGISTRATION NUMBER: 36,300
; REFERENCE/DOCKET NUMBER: MOBT:018 (38-2(10694))A
; NAME: Barbara S. Kitchell
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: MOBT:018 (38-2(10694))A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2030 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-705-937-7

Query Match 2.0%; Score 46.8; DB 2; Length 2030;
Best Local Similarity 52.0%; Pred. No. 0.12;
Matches 105; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1282 atcattcactatcatcaaacatttcacgtaagataaaaaaacattcccccacaaatgaga 1341
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 AGCGATAATAAATAGAGATAATAGTAAATAATAATAATAATAATAATAATAATAATAATAATA 767
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1342 gagacactacatctctcttattattataaataaagataaagataaagataaagataaagata 1401
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 707
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1402 aacacatatttgataaaattttactaaacatttttcttagtacttgtaataatcatgtct 1461
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
? TELEFAX: 215-568-3439
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5852 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 2378..5038
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 2378..5038
? US-07-867-106-2

Query Match      1.9%; Score 43.8; DB 1; Length 5852;
Best Local Similarity 46.8%; Pred. No. 0.74;
Matches 171; Conservative 0; Mismatches 192; Indels 2; Gaps 1;

QY 1921 ttcaaaactaaacaaacaaacttagaataatttggtaattatagcacacaattttcaaaaa 1980
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5312 TTTAAAAAAATTTTAAACATGGAATATATAGATCGATAGATCACTAATTTTAA 5371

QY 1981 tatcctagtcctcaaccactcaataattccaaatcccttgcaaaacacacaca 2040
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5372 AATTAAATATATTAAATTTATAAAAATTCAGTTCATCAAGATATATAGATAATTATTTA 5431

QY 2041 accttagaaactttgatttaatactataaaagcaataatgatatactaaacaatat 2100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5432 ATTATTTGAATTTTAAAAAATAAAAAAAAAAAAAAAATCAATATGTTTATGT 5491

QY 2101 caccatatatgatatataatgatgcagcaatacacacttaatttggtaaagcattaa 2160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5492 TTTAGATTTTAAATCTCGTCAATCATTTTAAATTAATAAATCGATACATAA--TTTTAA 5549

QY 2161 agcgagacaactctataacaccggttaattcaacaaccgttggtgcgagttcatgttt 2220
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5550 AAAAAACCCCTTTACATTTTATTTTAAATCCAAATTTATACATTTTATTTT 5609

QY 2221 ctccaactctttccctttccttacttatttatttctcacttaocttttctacta 2280
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5610 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 5669

QY 2281 atata 2285
   || | |
Db 5670 ATTTA 5674
```

Search completed: April 3, 2002, 05:16:12
Job time: 3982 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 04:39:16 ; Search time 2120.21 Seconds
(without alignments)
11783.708 Million cell updates/sec

Title: US-09-700-187-3

Perfect score: 2325

Sequence: 1 aagctttaaggcaaggaa.....ttttgatcttgacaagaaa 2325

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estfun.*
2: em_esthum.*
3: em_estin.*
4: em_estom.*
5: em_estpl.*
6: em_estba.*
7: em_estro.*
8: em_estov.*
9: em_htc.*
10: gb_estl.*
11: gb_est2.*
12: gb_htc.*
13: gb_gss.*
14: em_gss_fun.*
15: em_gss_hum.*
16: em_gss_inv.*
17: em_gss_pln.*
18: em_gss_pro.*
19: em_gss_rod.*
20: em_gss_vrt.*
21: em_gss_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	91.6	3.9	1101	13	CNS0039G
2	91	3.9	1101	13	CNS00L2
3	79.8	3.4	1101	13	CNS00EVL
4	75.6	3.3	1101	13	CNS0021J
5	74	3.2	1101	13	CNS0182P
6	72.4	3.1	1101	13	CNS00EVL
7	71.4	3.1	587	10	AW774987
8	70.8	3.0	1101	13	CNS006TE
9	70.6	3.0	1125	10	AL547503
10	70	3.0	928	13	CNS00DKY
11	70	3.0	1101	13	CNS00E07
12	70	3.0	1101	13	CNS00EPO

13	70	3.0	1204	13	CNS016E2
14	69.4	3.0	1061	13	CNS015LM
15	69	3.0	759	13	CNS06QXV
16	68.2	2.9	1101	13	CNS00CYH
17	67.4	2.9	928	13	CNS00DKY
18	67.2	2.9	1101	13	CNS000D1
19	67	2.9	1101	13	CNS0039G
20	66.4	2.9	694	13	A0853360
21	66.2	2.8	963	13	CNS00A4L
22	65.8	2.8	1101	13	CNS002FG
23	65.8	2.8	1101	13	CNS016LI
24	64.6	2.8	734	13	CNS010MP
25	64.6	2.8	1131	13	CNS034FO
26	64.4	2.8	1101	13	CNS00399
27	64.2	2.8	1146	13	CNS021G2
28	64	2.8	1201	13	CNS0162X
29	63.8	2.7	1101	13	CNS00EJ4
30	63.8	2.7	1101	13	CNS0100X
31	63.6	2.7	1190	13	CNS020N7
32	63.4	2.7	693	10	AV682300
33	63.4	2.7	893	13	CNS013XE
34	63.4	2.7	1101	13	CNS00EPO
35	63.2	2.7	1101	13	CNS017KE
36	62.8	2.7	609	13	CNS025K2
37	62.8	2.7	987	13	CNS014PQ
38	62.6	2.7	942	13	CNS018GS
39	62.6	2.7	1201	13	CNS0165X
40	62.6	2.7	1225	13	CNS0161D
41	62.2	2.7	643	13	B11768
42	62	2.7	1201	13	CNS0167M
43	61.8	2.7	859	13	CNS004YY
44	61.8	2.7	889	13	CNS006W4
45	61.6	2.6	963	13	CNS0075X

ALIGNMENTS

RESULT 1

CNS0039G

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS0039G 1101 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL063921 GI:4941778
GSS.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Oseogawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

[illegible]


```

RESULT      8
CNS006TE   1101 bp      DNA      GSS      03-JUN-1999
LOCUS       Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR14L09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065901
VERSION     AL065901.1 GI:4944869
KEYWORDS    GSS.
SOURCE      fruit fly,
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
            Genoscope.
            Direct Submission
            Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mannosier in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library's
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             location/qualifiers
            source          1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_lib="RPCI-98"
                        /clone="BACR14L09"
                        /notes="end : TET3"
BASE COUNT      417 a   102 c   21 g   232 t   329 others
ORIGIN
Query Match      3.08; Score 70.8; DB 13; Length 1101;
Best Local Similarity 30.1%; Pred. No. 0.078;
Matches 205; Conservative 128; Mismatches 341; Indels 7; Gaps 1;
QY 1291 actatcaaacatttcagtaaagataaaataattccaccacaaagagagagacacta 1350
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 400 ANNNNTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 459
QY 1351 catctctctattattataataaaatgtaaagaaaaatatagtaaaagtaacacatat 1410
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 460 TATCTTCTTAATATTTNCCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 519
QY 1411 ttgataaattataactaaactatttcttagtactgttaactgtaactgctgagggatttt 1470
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 520 AATATANTNNNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 579
QY 1471 acagtaaaagaacagggtagcccaacaaaagtgataattgtggagggtgcatctt 1530
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 580 AAAANAGWNTNRTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 639
QY 1531 tgtcggtgcaaaaaatgaaaccccaactgtgatatgtgtgcgactgctcgctgctac 1590
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 640 BMTAKGTSHGKWKVTMBKBTBMDKAKWNTDNNSTMTNSTKWTYTTNNMYWTHHTTATRT 699
QY 1591 attgaataatgaatgtctcttttataacgctttgtcctatgcgcg-----tattaccoat 1643
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 700 KTTAYATHTYATKTTTCTTCTTMTATBTTTWTATYATYATYATYATYATYATYATYATY 759

```

BASE COUNT	262 a	70 c	84 g	321 t	191 others							
ORIGIN												
Query Match	3.0%; Score 70; DB 13; Length 928;											
Best Local Similarity	33.5%; Pred. No. 0.11;											
Matches 120; Conservative	78; Mismatches 160; Indels 0; Caps 0;											
QY	1941	aactagaattttgtaattatagacaatttccaataatccttagcttcacacct	2000									
DB	571	wawttttaattwtatwttatwtttaaawttawwtatttatwtawaatatwwaaatatwttwata	630									
QY	2001	caataattccaaattcccgaatcccttgcaaacaccacacctctagaacctgtgatga	2060									
DB	631	ttaatattwtttwttawtttttatataaaaaawaaaawaaaataaaaaaaaaataawaa	690									
QY	2061	ataatctaataaaaggcaaatatgatcatctaaacaatatcacccatatatgtatgat	2120									
DB	691	wtAAAAAaaaaatttawtt	750									
QY	2121	aatatgatgcagcaacaatacttaatttgttaaagcataaagcgagacacctattaac	2180									
DB	751	wdGKNWNNNAWWWWWAAAWAAAANWWWWWWAAAAAANAANAANAANAANAANAANA	810									
QY	2181	accggtaattcaaacaocgtgttgctgcagtctatgtttcttcaccaactctttcccttt	2240									
DB	811	wdddddDWDKWAACKKKKKKKKKKKKKKKKKKKTKTCTTTKGARWTTTTTTTTTTTT	870									
QY	2241	ccttactattattttctoctacttacotctttctactactataatactatctctct	2298									
DB	871	TTT	928									
RESULT 11												
CNS00E07/c												
LOCUS	CNS00E07	1101 bp	DNA	GSS	04 JUN-1999							
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.											
ACCESSION	AL069440											
VERSION	AL069440.1	GI:4949583										
KEYWORDS	GSS.											
SOURCE	fruit fly. Drosophila melanogaster											
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.											
REFERENCE	1 (bases 1 to 1101) Genoscope.											
AUTHORS	Direct Submission											
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :											
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)											
COMMENT	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm. Location/Qualifiers 1. .1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98"											

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osodegawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial *EcoRI* digestion of Drosophila DNA provided by the BDGP from the isogenic strain y²; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29P01"
/note="end : T7"
289 a 155 c 118 g 307 t 232 others

```

ch 3.0%; Score 70; DB 13; Length 1101;
l Similarity 41.3%; Pred. No. 0.1;
157: Conservative 48; Mismatches 173; Indels 2; Gaps 1;

tatttttgcattttcaaaaaactaaacaaactagaatatatttggtaattatagca 1967

TTTTGGTTTNAATTAATAAAWAAAAATWTAWATAAATTAATWTAAATTT 542
 aaattttcaaaaatacctagtcttcacccactcaataattccaaatttccaaatccctt 2077

ATATAAAGAAATTTCAWTAATTATTTAAWTTTAAAT 600
cagaacatcacaccctctagaaacttggataataatctaataaagcaataatatgat 2087

TAATHYTAAGATTTTWTATATTTTATAATAAATAWTATATTATTAAATWHTT 660

WTWATAAAAAATWATATATATAATWATWATAATTATATYTYAMTAWMTTTTTTTW 720

TTTATCWAAGAAWAAWTWAAATTTATTATATAAAAAATWTTTTTAAATATATTTW 780

agtttcagtgttcttccaaactcttttccctttcctttactttattttattttctctactt 2267
|::|||:|:|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
TTTTAYHWYTTTWTWTAATTAATAAATAAAAWNTTYTTTTAAATTATATWWCTAH 840

cccttttttactaatatata 2287
 ||| | : ||| :
 TTATTACATAWAAATCTM 860

CNS016E2 1204 bp DNA GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.
 AL106628
 AL106628.1 GI:5623852

fruit fly.
GSS.
Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

/clone="BACR29p01"
/note="end : TET3"
366 a      66 c    104 g      351 t     214 others
ORIGIN
Query Match          3.0%; Score 70; DB 13; Length 1101;
Best Local Similarity 36.7%; Pred. NO. 0.1;
Matches 232; Conservative 82; Mismatches 314; Indels 5; Gaps 2;
1652 tagaatgggacaaatgaatttaatatatctgtcatgtgtgggtcaatttaattg 1711
   |:|::||::||::||::||::||::||::||::||::||::||::||::||::||::
1028 TWMAWYACAMWAYTWTATTATATTTTTTTTAAATATCWATATTTTWAATACAWA 969
   ||::||::||::||::||::||::||::||::||::||::||::||::||::
1712 tatcgtaaatggtaggacaactcatgcatacaattatcatactcagtgtcaatcacgt 1771
   |:|::||::||::||::||::||::||::||::||::||::||::||::||::
968 YATTTTWTATACATAATWWTTTTTTTATATACAATTTWAAAAATAAAAATTAACWAAATTTA 909
   |:|::||::||::||::||::||::||::||::||::||::||::||::||::

```

1772 gcaatgtgtttcttcccatgaattcacattgctaagaataa-ccaccttaaa 1830
 | : | | | | : | : | : | | | : | | | : | |
908 WAAACATTTWTTTAAAWTAATTWTACAWTTTAAAAAATWTTAAAWTAAA 849
 | : | | | | : | : | : | | | : | | | : | |
1831 tgttatccttcgcacacattccatccaattttaaacatttcaccattggaaaaca 1890
 | : | | | | : | | | : | | | : | | | : | |
848 WAAAAARWAATAAAATTTAYATWATATWTTAAAWWTATAWATTATMMWAATWTTATANNA 789

788 TTTT...TTTAAAWTWTWATWATTAATTTAAWTWWTTAATAAAAAATTTATT 729

1951 attttgtaattatagacaaattttcaaaaaatatcctaactctcaaccactcaataattca 2010

728 TATTTWTATWAAWTTTWTTTTWAATTTWYTTTAAATWTTAAAWTTTAAAWWTA 669

2011 caattcccaaaatccctccgcaaa -- acatcacacacccctagaaacccctggatccacacacac 2060

668 awtttaaaataaattwtatttaaaatwtgaawtataaaatttaawadwtatatawwtttttttwwc 609

2067 taataaaagcaataatgatatctctaaaacaatacaccatatatgttgatataaatatg 2126
 : :||| | ||| | | | | | | | | | | | | | | | | | | | | | | | |
608 WTTWAAATATATAAAAATWAAAAATTAAATTTTWTATTTTATATWWATAAAAAAMATWWWTTATT 549

2127 atcgagcaatacacacattaatttggtaagcattaaagcgagacaactctattaacaccggt 2186
 548 TATAAAWAAATATTATAWAAWTTGATTTTATTATTTTAAWTTTATTATATATWA 489

[illegible]

U
Y

486 TTATTAATWTTAAWTATTTTTTATTATACGCTTCMAAWTATTTMCCCCITTTTWTIM 429

2247 ctttattttatctctacttaccttttctact 2279
| | | | | : | | | | |
| | | | | | | | | | |

428 ATGTTTATTTTTHCTCTTNCNTTNTTNTT 396

Accession	Organism	Gene	Size	Accession
U00096	<i>Escherichia coli</i>	oriC	2454 bp	01-JUN-1999
U00097	<i>Escherichia coli</i>	oriT	2454 bp	01-JUN-1999
U00098	<i>Escherichia coli</i>	oriV	2454 bp	01-JUN-1999
U00099	<i>Escherichia coli</i>	oriW	2454 bp	01-JUN-1999
U00100	<i>Escherichia coli</i>	oriX	2454 bp	01-JUN-1999
U00101	<i>Escherichia coli</i>	oriY	2454 bp	01-JUN-1999
U00102	<i>Escherichia coli</i>	oriZ	2454 bp	01-JUN-1999
U00103	<i>Escherichia coli</i>	oriA	2454 bp	01-JUN-1999
U00104	<i>Escherichia coli</i>	oriB	2454 bp	01-JUN-1999
U00105	<i>Escherichia coli</i>	oriC	2454 bp	01-JUN-1999
U00106	<i>Escherichia coli</i>	oriD	2454 bp	01-JUN-1999
U00107	<i>Escherichia coli</i>	oriE	2454 bp	01-JUN-1999
U00108	<i>Escherichia coli</i>	oriF	2454 bp	01-JUN-1999
U00109	<i>Escherichia coli</i>	oriG	2454 bp	01-JUN-1999
U00110	<i>Escherichia coli</i>	oriH	2454 bp	01-JUN-1999
U00111	<i>Escherichia coli</i>	oriI	2454 bp	01-JUN-1999
U00112	<i>Escherichia coli</i>	oriJ	2454 bp	01-JUN-1999
U00113	<i>Escherichia coli</i>	oriK	2454 bp	01-JUN-1999
U00114	<i>Escherichia coli</i>	oriL	2454 bp	01-JUN-1999
U00115	<i>Escherichia coli</i>	oriM	2454 bp	01-JUN-1999
U00116	<i>Escherichia coli</i>	oriN	2454 bp	01-JUN-1999
U00117	<i>Escherichia coli</i>	oriO	2454 bp	01-JUN-1999
U00118	<i>Escherichia coli</i>	oriP	2454 bp	01-JUN-1999
U00119	<i>Escherichia coli</i>	oriQ	2454 bp	01-JUN-1999
U00120	<i>Escherichia coli</i>	oriR	2454 bp	01-JUN-1999
U00121	<i>Escherichia coli</i>	oriS	2454 bp	01-JUN-1999
U00122	<i>Escherichia coli</i>	oriT	2454 bp	01-JUN-1999
U00123	<i>Escherichia coli</i>	oriU	2454 bp	01-JUN-1999
U00124	<i>Escherichia coli</i>	oriV	2454 bp	01-JUN-1999
U00125	<i>Escherichia coli</i>	oriW	2454 bp	01-JUN-1999
U00126	<i>Escherichia coli</i>	oriX	2454 bp	01-JUN-1999
U00127	<i>Escherichia coli</i>	oriY	2454 bp	01-JUN-1999
U00128	<i>Escherichia coli</i>	oriZ	2454 bp	01-JUN-1999
U00129	<i>Escherichia coli</i>	oriA	2454 bp	01-JUN-1999
U00130	<i>Escherichia coli</i>	oriB	2454 bp	01-JUN-1999
U00131	<i>Escherichia coli</i>	oriC	2454 bp	01-JUN-1999
U00132	<i>Escherichia coli</i>	oriD	2454 bp	01-JUN-1999
U00133	<i>Escherichia coli</i>	oriE	2454 bp	01-JUN-1999
U00134	<i>Escherichia coli</i>	oriF	2454 bp	01-JUN-1999
U00135	<i>Escherichia coli</i>	oriG	2454 bp	01-JUN-1999
U00136	<i>Escherichia coli</i>	oriH	2454 bp	01-JUN-1999
U00137	<i>Escherichia coli</i>	oriI	2454 bp	01-JUN-1999
U00138	<i>Escherichia coli</i>	oriJ	2454 bp	01-JUN-1999
U00139	<i>Escherichia coli</i>	oriK	2454 bp	01-JUN-1999
U00140	<i>Escherichia coli</i>	oriL	2454 bp	01-JUN-1999
U00141	<i>Escherichia coli</i>	oriM	2454 bp	01-JUN-1999
U00142	<i>Escherichia coli</i>	oriN	2454 bp	01-JUN-1999
U00143	<i>Escherichia coli</i>	oriO	2454 bp	01-JUN-1999
U00144	<i>Escherichia coli</i>	oriP	2454 bp	01-JUN-1999
U00145	<i>Escherichia coli</i>	oriQ	2454 bp	01-JUN-1999
U00146	<i>Escherichia coli</i>	oriR	2454 bp	01-JUN-1999
U00147	<i>Escherichia coli</i>	oriS	2454 bp	01-JUN-1999
U00148	<i>Escherichia coli</i>	oriT	2454 bp	01-JUN-1999
U00149	<i>Escherichia coli</i>	oriU	2454 bp	01-JUN-1999
U00150	<i>Escherichia coli</i>	oriV	2454 bp	01-JUN-1999
U00151	<i>Escherichia coli</i>	oriW	2454 bp	01-JUN-1999
U00152	<i>Escherichia coli</i>	oriX	2454 bp	01-JUN-1999
U00153	<i>Escherichia coli</i>	oriY	2454 bp	01-JUN-1999
U00154	<i>Escherichia coli</i>	oriZ	2454 bp	01-JUN-1999
U00155	<i>Escherichia</i>			

BACR29P01 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.
 AL069493
 ACCESSION
 AT069403.1 GI:1404626

ORIGIN	AL0009435.1	GI:4943030
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	

Eukaryota; Metazoa; Artthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1101)		Genoscope.	
		Direct Submission	
		Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage :

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2002, 03:52:00 ; Search time 2120.21 Seconds
(without alignments)
60.819 Million cell updates/sec

Title: US-09-700-187-1

Perfect score: 12

Sequence: 1 ggattttacagt 12

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun:.*
2: em_esthum:.*
3: em_estin:.*
4: em_estom:.*
5: em_estpl:.*
6: em_estba:.*
7: em_estov:.*
8: em_estov:.*
9: em_hc:.*
10: gb_est1:.*
11: gb_est2:.*
12: gb_hc:.*
13: gb_gss:.*
14: em_gss_fun:.*
15: em_gss_hum:.*
16: em_gss_inv:.*
17: em_gss_pln:.*
18: em_gss_pro:.*
19: em_gss_rod:.*
20: em_gss_vrt:.*
21: em_gss_other:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	88	13	AZ925237 4910 ez32
2	12	100.0	118	10	AW217731 EST296445
3	12	100.0	126	13	AZ289196 RPCI-23-5
4	12	100.0	130	10	AI610647 tp20d02.x
5	12	100.0	147	10	BE145737 IL5-HT020
6	12	100.0	147	11	D25785 HUMGS04153
7	12	100.0	150	10	AA579315 nf36e06.s
8	12	100.0	152	10	AU180504 AU180504
9	12	100.0	155	13	AQ009485 CIT-HSP-2
10	12	100.0	165	10	AV333933 AV333933
11	12	100.0	165	11	BF881282 QV1-ET018
12	12	100.0	171	11	BF909645 PM3-UT005

c 13	12	100.0	172	10	AI029683 UI-R-C0-1
c 14	12	100.0	173	13	AZ818519 2M0088A10
c 15	12	100.0	177	13	AQ985182 RPCI-23-3
c 16	12	100.0	180	11	BF235518 602025332
c 17	12	100.0	184	10	AV054690 AV054690
c 18	12	100.0	185	10	AV071084 AV071084
c 19	12	100.0	185	10	BB357868 BB357868
c 20	12	100.0	185	11	BG942699 ax28e01.x
c 21	12	100.0	186	10	AI345298 t868c12.x
c 22	12	100.0	189	11	F29452 HSPD19337.H
c 23	12	100.0	190	10	AV288772 AV288772
c 24	12	100.0	193	10	AV285178 AV285178
c 25	12	100.0	194	10	AI136924 UI-R-C2P-
c 26	12	100.0	202	11	D60959 HUM143A06B
c 27	12	100.0	207	10	AI572468 tp23g01.x
c 28	12	100.0	210	10	AV336005 AV336005
c 29	12	100.0	216	13	B25087 F24122TR IG
c 30	12	100.0	217	10	AI563911 tn31c07.x
c 31	12	100.0	220	13	AQ278868 CITBI-E1-
c 32	12	100.0	223	10	BE236288 144014 MA
c 33	12	100.0	223	10	BE236293 144020 MA
c 34	12	100.0	224	10	AV252038 AV252038
c 35	12	100.0	224	10	BB512144 BB512144
c 36	12	100.0	225	10	BB256670 BB256670
c 37	12	100.0	226	11	BI220740 602938853
c 38	12	100.0	227	10	AA129692 z091e04.S
c 39	12	100.0	227	11	BI293223 UI-R-DK0-
c 40	12	100.0	228	10	BB171532 BB171532
c 41	12	100.0	228	10	BB310621 BB310621
c 42	12	100.0	230	10	BB152264 BB152264
c 43	12	100.0	231	10	AV368688 AV368688
c 44	12	100.0	231	10	BB503511 BB503511
c 45	12	100.0	231	10	BB510596 BB510596

ALIGNMENTS

RESULT 1
AZ925237/c
LOCUS 88 bp DNA
DEFINITION 4910.ez32kl8.sl Saccharomyces paradoxus N17 Saccharomyces paradoxus genomic clone 4910.ez32kl8.sl, DNA sequence.
ACCESSION AZ925237 GI:13496136
VERSION 1
KEYWORDS GSS.
SOURCE Saccharomyces paradoxus.
ORGANISM Saccharomyces paradoxus.
REFERENCE 1 (bases 1 to 88)
AUTHORS Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish ,W.R., Waterston,R.H. and Johnston,M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mjgenetics.wustl.edu
Class: random plasmid subclone.
Location/Qualifiers
1. .88
/organism="Saccharomyces paradoxus"
/strain="N17"
/db_xref="taxon:27291"
/clone="4910.ez32kl8.sl"
/clone_lib="Saccharomyces paradoxus N17"
/note="Random genomic sequence"

BASE COUNT 26 a 12 c 8 g 42 t

THIS PAGE BLANK (USPTO)

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 88;
 Best Local Similarity 100.0%; Pred. No. 7.3e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||||
 Db 40 GGATTTTACAGT 29

RESULT 2

AW217731/c
 LOCUS
 DEFINITION EST296445 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone cTOG6E5, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

1 (bases 1 to 118)

REFERENCE
 AUTHORS van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.
 Generation of ESTs from tomato flower tissue
 Unpublished (1999)

- TITLE

JOURNAL

COMMENT

Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>

FEATURES

source

Location/Qualifiers

1..118
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cTOG6E5"
 /clone_lib="tomato flower buds 8 mm to pre-anthesis,
 Cornell University"
 /tissue_type="flower"
 /dev_stage="buds 8mm-to-preanthesis"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Tanksley; Flower buds and flowers were
 taken from greenhouse plants (4-8 wks old, TA496). They
 were immediately frozen in liquid nitrogen and then
 size-separated while remaining frozen."

BASE COUNT

ORIGIN

49 a 21 c 16 g 32 t

Query Match

Best Local Similarity 100.0%; Score 12; DB 10; Length 118;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12

|||||

Db 47 GGATTTTACAGT 36

RESULT 3

AZ289196
 LOCUS
 DEFINITION RPCI-23-59B23.TJB RPCI-23 Mus musculus genomic clone RPCI-23-59B23,
 DNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 130)
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AZ289196.1 GI:9530982

GSS.

house mouse.

Mus musculus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 126)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Aknret,

B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 59 row: B column: 23

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..126

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-59B23"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

38 a 20 c 30 g 38 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 126;

Best Local Similarity 100.0%; Pred. No. 7.1e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12

|||||

Db 93 GGATTTTACAGT 104

RESULT 4

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Qy

Db

THIS PAGE BLANK (USPTO)

JOURNAL COMMENT

Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: crapsb@mail.nih.gov
 Tissue Procurement: Christopher Mokaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 cDNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Seq primer: -40up from Gibco
 High quality sequence stop: 1.

FEATURES

source
 1. .130
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:2188323"
 /clone_lib="NCL_CGAP_Gas4"
 /tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"
 /lab_host="DH10B"
 /note="Organ: Stomach; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.69 kb. Life Technologies catalog #: 115497011"

BASE COUNT 36 a 21 c 28 g 45 t
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 130;
 Best Local Similarity 100.0%; Pred. No. 7.1e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||
 Db 115 GGATTTTACAGT 104

RESULT 5
 BE145737/c
 LOCUS BE145737 147 bp mRNA EST 21-JUN-2000
 DEFINITION IL5-HT0207-231099-006-A01 HT0207 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE145737
 VERSION BE145737.1 GI:8608461
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 147)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,C.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-IL5-HT0207-231
 099-006-A01&t3=1999-10-23&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 56
 High quality sequence stop: 147.
 Location/Qualifiers

source
 1. .147
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0207"
 /dev_stage="Adult"
 /note="Organ: head neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 31 a 35 c 12 g 69 t
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 147;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||
 Db 132 GGATTTTACAGT 121

RESULT 6

D25785/c
 LOCUS D25785 147 bp mRNA EST 30-NOV-1995
 DEFINITION HUMG504153 Human colon mucosa Homo sapiens cDNA clone cml984 3', mRNA sequence.

ACCESSION D25785
 VERSION D25785.1 GI:500469
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 147)

Okubo,K., Yoshii,J., Yokouchi,H., Kameyama,M. and Matsubara,K.
 Global analysis of gene expression in colon mucosa: a large scale random cDNA sequencing analysis
 Unpublished (1994)
 Contact: Okubo,K., Itoh,K., Yoshii,J., Yokouchi,H. and Matsubara,K.
 Institute for Molecular and Cellular Biology
 Osaka University
 3-1 Yamada-oka,Suita,Osaka 565,Japan.

Location/Qualifiers
 source
 1. .147
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="cml984"
 /clone_lib="Human colon mucosa"
 /note="Adult male, tissue_type = colon mucosa"

BASE COUNT 63 a 24 c 34 g 26 t
 ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 147;
 Best Local Similarity 100.0%; Pred. No. 7e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggattttacagt 12
 |||
 Db 57 GGATTTTACAGT 46

THIS PAGE BLANK (USPTO)

```

ORGANISM      Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 152)
Sanaka,E., Hori,H., Naruse,K., Mitani,H. and Tanaka,M.
Medaka EST analysis
Unpublished (2001)
Contact: Emi Sanaka
Department of Biological Sciences
Graduate School of Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
Tel: 81-52-789-2973
Fax: 81-52-789-2974
Email: sanaka@bio.nagoya-u.ac.jp
This clone was isolated from Medaka eye cDNA library (SNK01) 5'end.
Location/Qualifiers
1..152
/organism="Oryzias latipes"
/strain="wild type"
/db_xref="taxon:8090"
/clone="NCY10.02c"
/clone_lib="Medaka eye cDNA library (SNK01)"
/tissue_type="eye"
/dev_stage="adult"
/note="Wild samples from Okayama Pref.(Southern part of
Japan)"
BASE COUNT      56 a      23 c      30 g      43 t
ORIGIN
Query Match      100.0%; Score 12; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      .1 ggattttacagt 12
        |||||
Db       81 GGATTTTACAGT 92

RESULT          9
AQQ009485
LOCUS          27-JUN-1998
DEFINITION    CIT-HSP-2283N16.TRB CIT-HSP Homo sapiens genomic clone 2283N16, DNA
sequence.
ACCESSION     AQQ009485
VERSION       AQQ009485.1 GI:3128870
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other_GSSs:   CIR-HSP-2283N16.TFB
Contact:      Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

```

THIS PAGE BLANK (USPTO)

Tel: +55-11-27049222

THIS PAGE BLANK (USPTO)


```

Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV1st2-QV1-ET0183-
01200-529-d05_1st3=2000-12-02&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 165.
Location/Qualifiers
1..165
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0183"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
34 a 46 c 35 g 50 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 165;
-Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggattttacagt 12
|||||
Db 96 GGATTTTACAGT 107

RESULT 12
BF909645 171 bp mRNA EST 18-JAN-2001
LOCUS PM3-UT0058-181000-007-g01 UT0058 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF909645
ACCESSION BF909645
VERSION BF909645.1 GI:12301103
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 171)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202863
MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3st2-PM3-UT0058-
181000-007-g01&t3=2000-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 113.
Location/Qualifiers
1..171
FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT0058"
/dev_stage="Adult"
/note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
34 a 42 c 49 g 46 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 12; DB 11; Length 171;
-Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ugattttacagt 12
|||||
Db 117 GGATTTTACAGT 128

RESULT 13
AT029683 172 bp mRNA EST 04-JUL-1999
LOCUS UI-R-CO-iy-b-01-0-UI.s1 UI-R-CO Rattus norvegicus cDNA, clone
DEFINITION UI-R-CO-iy-b-01-0-UI 3', mRNA sequence.
ACCESSION AT029683
VERSION AT029683.1 GI:4300245
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 172)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
MEDLINE
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247509.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult liver library. cDNA Library Preparation: M. Fatima Bonaldi,
Ph.D. Clone distribution: clones will be available through Research
Genetics. This clone is also available through the I.M.A.G.E.
Consortium at LNL (info@image.llnl.gov). IMAGE ID=1782730
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1..172
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CO-iy-b-01-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CO
library is a subtracted library derived from the UI-R-A1
and UI-R-E1 libraries. The UI-R-A1 library consisted of a
mixture of individually tagged normalized libraries

```

THIS PAGE BLANK (USPTO)

constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-EI library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-Al and UI-R-EI clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-Al and UI-R-EI library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)*

BASE COUNT 32 a 47 c 32 g 61 t
ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 172;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 29 GGATTTTACAGT 18

RESULT 14

AZ818519 173 bp DNA GSS 20-FEB-2001
LOCUS
DEFINITION 2M0088A10R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0088A10 R, DNA sequence.

ACCESSION AZ818519
VERSION AZ818519.1 GI:12988427
KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 173)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.

REFERENCE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dduenn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0088 row: A column: 10

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 173.

Location/Qualifiers

1. 173

FEATURES

source

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0088A10"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 41 a 41 c 46 g 45 t
ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 173;
Best Local Similarity 100.0%; Pred. No. 7e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggattttacagt 12
|||||
Db 134 GGATTTTACAGT 145

RESULT 15

AQ985182 177 bp DNA GSS 30-JAN-2000
LOCUS
DEFINITION RPCI-23-307M5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-307M5,
DNA sequence.

ACCESSION AQ985182
VERSION AQ985182.1 GI:6818387
KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 177)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.

REFERENCE Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

JOURNAL

COMMENT Other_GSSs: RPCI-23-307M5.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 307 row: M column: 5

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 177

/organism="Mus musculus"

/strain="C57BL/6J"

THIS PAGE BLANK (USPTO)

```
/db_xref="taxon:10090"  
/clone="RPCI-23-307M5"  
/clone_lib="RPCI-23"  
/sex="Female"  
/lab_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:  
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT      56 a      48 c      20 g      52 t      1 others  
ORIGIN
```

```
Query Match      100.0%; Score 12; DB 13; Length 177;  
Best Local Similarity 100.0%; Pred. No. 7e+03;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY      1 ggattttacagt 12  
        ||| ||||| |||||  
Db      139 GGATTTTACAGT 128
```

```
Search completed: April 3, 2002, 04:39:12  
Job time: 2832 sec
```

THIS PAGE BLANK (USPTO)